

标准化调查访问：
如何实现访问员相关误差最小化

STANDARDIZED SURVEY INTERVIEWING：
MINIMIZING INTERVIEWER-RELATED ERROR

弗洛伊德·福勒（Floyd J.Fowler,Jr.）
托马斯·曼吉奥诺（Thomas W.Mangione） 著

孙 龙 徐方敏 译

张小劲 校

重庆大学出版社

作译者简介

弗洛伊德·福勒(Floyd J. Fowler, Jr.)

1966年在密歇根大学获得社会心理学博士学位。福勒博士的职业生涯即致力于调查方法的研究、改进和应用。他是多项调查研究课题的主持人,这些课题涉及地方人口趋势、公众对地方政府和服务的态度、博彩法的执行情况、种族紧张、犯罪恐惧感、犹太人的认同以及老龄人口的需求等问题。福勒博士在哈佛公共卫生学院及其他机构讲授调查方法课程,并曾连续14年担任马萨诸塞大学波士顿分校问卷调查研究中心的主任。

托马斯·曼吉奥诺(Thomas W. Mangione)

1973年在密歇根大学获组织心理学博士学位。此后,他任职于马萨诸塞大学波士顿分校调查研究中心。他参与了100多项调查课题,涉及饮酒、吸毒、赌博、犯罪及犯罪恐惧感、安全带使用、工作压力、精神健康及艾滋病知识,以及导致艾滋病的风险的行为等众多主题。

曼吉奥诺博士现在是调查研究中心的高级研究员。在任职的16年中,他先后担任了调查研究中心的副主任和主任,同时还在哈佛大学和波士顿大学讲授调查方法课程。

张小劲

清华大学政治学系教授、主任,中国人民大学政治学博士,长期从事政治学方法论的教学与科研工作。

孙　龙

中国人民大学国际关系学院政治学讲师,2005年于北京大学获社会学博士学位,长期从事问卷调查和实证研究方面实践与研究工作。

徐方敏

2009年于中国人民大学获政治学硕士学位。

前言

在过去 35 年,自从 Herbert Hyman 及其同仁就问卷调查之中的访问员问题撰写了一本具有标志性意义的著作以来,社会科学家开始认识到,访员因素也是导致调查资料出现误差的原因。尽管自 1954 年以来,访员在方法论研究中从未成为“热门”话题,但是我们仍认为,应该对该领域日渐丰富的知识进行总结。

我们了解到,在问卷调查中,访员会影响他们所获得的答案,有一些可行的策略能够降低与访员相关的误差。但是,时至今日,这些策略在很大程度上尚未应用于调查实践。我们撰写此书的一个目的,即是在考虑成本一效益的情况下提出相关建议,从而更好地进行问卷调查;我们还将为那些正在寻求资助的调查者提供证据:相对较小的花费,完全可以对调查数据的质量造成很大的差别。

很显然,本书主要的读者群包括那些直接涉及总体调查(population surveys)的实施者。但是,访员为了很多不同的研究目的而搜集资料。任何人只要是依靠其他人提供信息并加以汇总列表,或转成量化的形式,就会对如何降低访员相关误差的方法感兴趣。因此,本书在写作之时,考虑了非常广泛的读者群,并不要求他们具有研究方法、统计学或者其他类型的特殊训练。

我们很严肃地认为,调查研究是一门社会科学。我们认为,社会科学中的测量,完全可以和其他科学一样严谨(rigorous)且有效(valid)。我们会害怕听到别人把访问或是问卷设计说成是一种艺术,而不是一种科学。为了提高测量在调查中的效度,有一套必须遵循的程序,这与其他科学中所必有的努力是一样的。本书将简要介绍这些程序,告诉读者问卷调查的访员是如何影响数据的,以及如何减少这些效应,从而产生更有效度的数据。

对于调查资料的使用者,调查中与人有关的部分总是会受到质

疑。有一些批评家认为,一致且标准化的访问是不可能的;另外一些调查数据的使用者则认为,标准化是给定的,是理所当然的。然而,这两种观点皆非尽然。通过访问来进行测量是可能的,但绝非易事。如果没有好的访问,就不可能做到——至少不会做得很好。访问员和受访者都是普通人,有效的测量程序必须考虑到这一现实。我们希望,在考虑到访问得以进行的现实条件下,我们能够就如何进行高质量的访问作出较好的阐述。

下文所述就是更为详细的内容。

弗洛伊德·福勒

托马斯·曼吉奥诺

致谢

写作本书的动力，以及书中所列的诸多数据，均来源于一个大型实验项目，该项目旨在探讨访问员培训和督导(supervision)的作用。项目获得了美国卫生服务研究中心(National Center for Health Services Research，NCHSR)的资助，项目号为Grant# 3-R18-HS04189。

感谢美国卫生服务研究中心提供的资助。卫生状况的研究者们投入越来越多的资源以收集调查数据，而只有调查(survey)才能提供他们所需要的信息。然而，评估和设计调查方法的花费却只占相当小的份额。在方法论研究方面仍需要持续的投资，我们很高兴美国卫生服务研究中心将本项研究作为整个卫生服务研究项目的一部分。

我们要感谢美国卫生服务研究中心的项目官员 Bill Kitching 和 Stephen Marcus，他们为本项研究投入了时间、精力和兴趣。

实施这一大型实验项目本身即是一个巨大的成就。在完成这一复杂工作的诸多人员之中，Barbara Thomas，Alice Fehlhaber，Ruth Pradise，Kim Streitburger，Dottie Cerankowski 扮演了非常关键的角色。

当时任职于哈佛公共卫生学院、现任教于明尼苏达大学的 Tom Louis 所提供的统计学咨询，也是一个非常重要的支持。

我们还要将特别的感激之情献给 Charles F. Cannell。在本书之中，连续引用了他的诸多研究，而他对本书的智识贡献远远超过了本书的这些引证。没有人会比他更多地增进了我们对访员的理解。

我们感谢 Anstis Benfield 和 Peggy Carter 打印了本书的手稿。

我们感谢许多同仁通过多种形式所给出的建议，以及由此对本书所作出的贡献！当然，本书所包括的资料和结论，均由作者独立承担责任。

目录

第1章 什么是标准化调查访问

What Is a Standardized Survey Interview?

本书是就如何提高调查数据质量的问题进行持续研究的成果。我们所给予特别关注的，是那些为了收集某一人群的量化数据的问卷调查。例如，估计拥有小汽车的人口的百分比，或者在过去一年之中曾经住院者所占的百分比。这些数据是通过向被调查者提问而获得的。虽然有些调查依赖于受访者自填问卷，即受访者通过填写问卷而书面回答问题，但本书所关注的则是那些通过访问员提问并记录答案的众多调查。本书的主要目的在于论述如何操作此类访问，从而最大程度地减少与访员有关的误差。

调查访问的性质

访问可以被称为“有目的的对话（conversation）”（如，Kan & Cannell，1958）。但是，访问具有多种类型，会谈的特定目的会影响到其所适用的程序。此外，如果你要界定什么是成功的访问，以什么标准来评判访问质量，那么，这些问题都会因目标的差异而有所不同。

标准化的调查访问（standardized survey）仅仅是众多访问类型之中的一种。与其他类型的访问相比，标准化的调查访问具有两个基本特点：

1. 对话由提问和回答两个基本要素组成；
2. 参与者界限明确，角色间互不重复：一个人提问（访问员），

另外的人回答问题（受访者）。

将某些特定的对话归类为访问，最关键的要素是不同角色之间存在着明确的分工。虽然访问员可能也会主动提供信息或解释，但这些行为仅仅是为提问作准备；同样，受访者有时也会向访问员提问，但这也是为了成功地完成答问。

调查访问的设计，要求提供某一人群的描述性统计信息；与出于其他原因的访问相比，问卷调查的访问存在如下几个方面的特点：

1. 受访者个体之所以引起研究者的兴趣，仅仅由于他们是将要被描述的某一总体（population）中的一员。一般而言，他们是该总体的代表性样本。无论他们是如何被选中的，这些个体的回答之所以有价值，是因为他们可以帮助研究者描述其所属的一个人群，而非研究者对这些个体的受访者本人的独特性感兴趣。
2. 调查结果将被用来对总体进行定量描述，或者进行数量化描述。至少，研究者会将受访者分为几个离散的、界限明确的类型，并估计每种类型在总体中所占的百分比。调查结果通常表述为如下形式："现在，有 4% 的劳动者失业。"人们也可以描述不同变量之间的关系，例如："对工作不满意者，其请病假的天数要多于那些对工作表示满意的工人。"
3. 受访者提供的答案即是进行测量的结果，这也就是人们所要分析的数据。受访者会给出特定类型的答案；而所谓描述，就是这些答案的分布状况的直接结果。

调查是否成功，其标准就是数据在多大程度上测量出了研究者所要描述的总体的某些方面。调查的目标在于提供精确的统计数据。

调查中误差的来源

调查是一个测量过程，包括一系列步骤。为了对以调查为基

础的估计质量进行评估,需要考虑调查过程的每一个方面:

1. 任何基于样本的调查估计,都会受制于抽样误差。这种误差来源于这么一个事实,即样本不会正好与该样本所在的总体在所有方面都完全吻合。抽样方式及样本规模,都会影响到样本对总体的代表性。此外,调查估计是以对问题的实际回答情况为基础的。被抽取的被调查者经常拒访,或者由于其他原因而没有接受访问。被抽取的被调查者进行回答的比率,以及他们的答案在多大程度上具有代表性,是评估调查数据的另外一个关键问题。
2. 在回答的精确性上,不同的问题存在非常大的差异。调查所要获取信息的类型,以及具体问题的措辞,都会直接影响回答的质量。
3. 用于收集数据的程序同样重要。数据收集是通过电话还是面对面接触,访谈的情境,访问员与受访者互动的方式,都会影响到调查误差的大小与类型。
4. 将调查的答案转换为数字形式以用于计算机分析,也可能导致误差。编码时,如果编码员对编码规则的应用方式不统一,或者判断失误,那么也会产生误差。此外,数据录入失误也会导致整个调查产生误差。

表 1.1 对调查中的这些误差来源进行了总结。

表 1.1 调查中的误差来源

A. 抽样导致的误差:

1. 抽样框(sampling frame)或抽样底册没有包括要描述的总体的所有元素,这样会使某些类型的人群被排除在外;
2. 刚好存在某种可能性,样本可能没有反映它所属的总体;
3. 被样本抽中的人有时会拒绝回答所有或部分问题。

B. 提问导致的误差:

1. 问题被误解;
2. 受访者无法提供或无法准确回忆问题所要求的信息;
3. 受访者不愿意准确回答。

续表

C. 访问员导致的误差:
1. 没有逐字逐句提问;
2. 进行诱导式提问(probe directively);
3. 向受访者询问的方式,导致有偏见的答案;
4. 对答案的记录不准确。
D. 数据处理过程导致的误差:
1. 编码员应用编码规则的方式不统一,或者在选择代码时判断失误;
2. 在转化为计算机可以利用的文件时录入失误。

访问员的作用

表 1.1 所列出的误差来源中,绝大部分不是访问员的责任。调查通常会由一个团队来实施。某个人或几个人(即"研究者")决定调查方案,包括调查目标、抽样设计、将要询问的具体问题,以及将要使用的数据采集程序。一般来说,数据采集工作由另外一组人员来实际执行。尽管在一些小型调查中,研究人员会亲自执行全部或部分的访问工作,和/或只使用一位访问员,但绝大部分问卷调查的重要特点是,研究者设计数据采集程序,而其他许多人来执行数据采集工作。

访问员确实会影响回答率(response rate),即样本之中同意接受访问者所占的百分比。但是,调查误差的主要来源在于,他们以影响受访者的方式来影响数据采集过程。更具体地说,访问员引起误差的原因在于没有被标准化。研究者与访问员进行合作的挑战在于,如何将访问过程标准化。

作为测量的调查:标准化的必要性

根据《韦伯斯特(Webster)大辞典》的说法(Guralk,1976),测量是指"根据某一标准来发现或者估计事物的范围(extent)、尺度(dimension),等等"。测量过程的关键在于标准化。在所有科学

中,有意义的测量在于将同一程序应用于各种不同的情况,这样的结果可以进行比较和解释,从而反映出被测对象的真实差异。问卷调查也是如此。在这里,标准化的测量过程是提出一个问题,"测量"结果即访问员将每一个被调查者的回答记录下来。标准化的目的在于让每个受访者都回答相同的问题,并以相同的方式进行记录,这样,答案的差异性就可以准确地解释为受访者之间的差异,而非生成答案过程中产生的差异。

访问应当标准化,但说起来很容易,做起来却很难。我们可以举例说明:

问题:"您如何评价您年龄最大的孩子就读的学校——非常好,比较好,一般,比较差?"

受访者:"这个很难说。我孩子现在上二年级。我认为他们在算术和阅读技能训练方面做得不好。但在这个年龄,他们所学内容并不是都那么重要。在那些方面,孩子们可以赶上来。另一方面,我认为他们在娱乐方面做得很好,比如学校的体育馆和游乐设施很好,给孩子们创造了很多机会一起玩耍。我认为,对这个年龄段的孩子来说,这点非常重要。"

分析:这位受访者对所询问的问题存在着质疑,这是合乎情理的。学校包含诸多方面的内涵,包括运动设施、教师、管理、课程以及其他学生。而所询问的这个问题强迫被调查者综合考虑一个学校的不同方面,并且以某种方式将这些方面综合到一起,最后给出一个评估。

访问员或者采取非标准化的方式进行回应,或者采取标准化的方式进行回应。

非标准化应对方式1　:"我们就学校进行提问时,我想教育因素应该是最重要的一个方面。我想您应当依据教育项目的情况来进行评价。"

非标准化应对方式2　:"我们请您将学校作为一个总体进行评估。您可以将您认为最有价值的方面给予最重要的权重。"

标准化应对:“显然,您对一个学校进行评估的时候会考虑很多因素。调查,就是人们提出相同的问题,让他们依据自己认为最合适的方式对每个问题进行理解和回答。现在我再念一遍问题,然后请您从您的角度选择一个最合适的答案。”

进一步研究:可以肯定,一部分受访者会依照非标准化应对方式1所建议的方式来回答问题,而其他的则会遵循非标准化应对方式2所给出的参考框架来进行回答。关键是,受访者对参考框架的选择会受到访问员的非标准化的追问和解释的影响。如果访问员都以标准化的方式进行提问,我们不知道受访者会有什么样的考虑(除非询问其他问题),但是我们知道,答案的差异源于受访者对问题的理解和回答不尽相同,而与访问员的行为无关。这才是我们所追求的。

受访者:“我想最接近的答案应该是,我认为学校在某些方面非常好,某些方面并不是很好。”

分析:可以说这位受访者提供了一个经过深思熟虑的精确答案。但是,我们不可能把这个回答同其他受访者的回答合在一起来进行定量描述。尽管有些调查要求受访者用自己的语言来回答问题(我们称之为开放性回答或描述性回答),但是,大部分问题都和这个问题一样,要求受访者在所给出的一系列答案中选择一个最能描述他们的感受或者情境的答案。将回答形式具体化是标准化的一部分。

访问员的标准化应对:“我们请您选择一个选项,这样我们可以将您的回答与其他人的回答进行比较。这些选项中,哪个最能代表您对这所学校的总体评分——非常好,比较好,一般,或者比较差?”

显然,接受一个符合情况的答案比较简单,相对而言,强迫受访者从一系列选项中选择一个可能并不很符合他或她感受的选项则比较困难。同样,访问员会非常容易而且自然地采用非标准化方式的解释,帮助受访者回答。但是,如果测量就是让不同的

访问变得统一起来，那么就必须采取标准化措施。这也就是我们所说的标准化访问的关键。

与其他两类访问的比较

为了把握标准化访谈的特殊要求，我们将标准化访问与如下两类访问进行比较：

A. 晚间脱口秀节目的访谈。

B. 医生向患者询问病史。

从表面上看，将脱口秀访谈与标准化访问进行比较确实很荒谬。但是，这种访谈也符合访问的两个基本标准，即对话主要包括提问和回答，参与者有明确分工，即一个人主要负责提问，另外一个人主要负责回答。

两者最根本区别在于，脱口秀节目中双方互动的目的在于娱乐。节目结束后答案不会被用于任何目的，甚至没人会将其记录下来。

脱口秀节目的访谈问题可能是有事先设计和结构化的，但是，它们与节目里其他嘉宾被问到的问题极少有关联，或没有任何关联。有些访谈者试图"调动"嘉宾，给他们机会充分表达自己，同时尽量不让自己的角色影响互动。虽然这也是我们所认为的"优秀"访问员的特征，但是有些主持人也会利用问题来取悦观众，有时会将嘉宾的话语断章取义。从娱乐价值来看，当主持人表现得像"优秀"访问员时，这一过程可能很成功。

最后，脱口秀访谈节目的焦点每次只有一个特殊的嘉宾，不能用于进行定量分析。我们也不能将一个嘉宾的回答与其他嘉宾的回答进行比较。访谈结果没有后续用途，除非重播或用于周年秀(anniversary shows)。

当医生向患者询问病史时，其互动结构与调查访问非常类似。谁提问，谁回答，毫不含糊。此外，医生实际上往往是按照同一张清单来询问所有病人过去的情况和就诊事件的。

但是，医患之间的问答与标准化访问还是有一些区别。首先，其目的在于正确把握某一特定患者的特征。人们没有兴趣将

其推广到其他患者,也不会将其与其他患者进行比较。因此,对每个患者的用词很可能各不相同。虽然医生开始可能对所有患者按照同一模式进行提问,但他们也往往会自由提问;而如果患者的回答看起来符合某种情况,则会就某些问题进行深入调查。

此外,多数医生并不关心提问是否具有诱导性。医生提示一些词汇来描述患者的感受,这样他们可以将病人的经验纳入医学框架之中,这种情况并非个别现象。病历记录下来的与其说是患者自己提供的词汇或者答案,不如说是医生依据病人的回答而得出的结论。这样,如果一个患者说他有"心脏问题",医生会考虑这一信息,但医学结论会以医生自己的判断为基础。大多数时候,病史询问的记录资料也是医生的结论,而非病人自己的回答。参见表1.2。

表1.2 各类访谈的比较

类型	量化目标	答案用途	问题标准化
脱口秀	无	娱乐	无
病史	将病人状况记录归档	为诊断提供信息	主题相同而问题不同
标准化调查	描述一个群体	转化为可供研究的数据	所有访问采用相同问题

实际上,在多数情况下,访问员遵循一些非诱导性的标准化原则,对于访问是有益的。如果主持人的访谈是以回答者为导向,那么,尽量避免主持人自己观点和偏好的干扰,脱口秀节目将更有趣味,也更为多样化。同样,医生也可以受益于标准化的访问方法。因为他们知道,他们所得出的结论来自于问诊的结果,而并非是病人自己口述病状的结果;医生也许会使用一些引导性的访问技术,这会使患者同意医生的描述但在事实上并没有完全捕获患者自己的体验。医生的结论是至关重要的,但是只有充分利用了病人所能提供的信息时,这些结论才最为准确而有效。

如果说访问是一类互动形式的话,那么,标准化访问调查实际上就是这类互动的一个特例。但是,标准化访问所面临的挑战,例如让受访者精确而完整地回答问题,使访问员对答案的影

响最小化等,可以应用于大多数访谈情境。通过特定的程序和技术将一项涉及众多访问员和诸多受访者的调查实现标准化,这将是一个特殊的挑战,但是,在标准化访问调查中发挥作用的这些程序和解决方案对其他类型的访问也是同样适用的。

什么时候有必要将社会研究标准化

在人们收集信息用于社会研究之际,并非任何时候都可以采用标准化访问。

探索性研究(exploratory research)就不宜使用标准化访问。在应用标准化访问的时候,人们仅仅需要获取所提问题的答案。而在研究的探索性阶段,主要目标在于提出什么样的问题。同样,标准化访问的一个优点是将答案限定并加以结构化,但是,研究者只有在答案选项是全面而恰当之时才能做到这点。

个案研究(case study),其目的在于充分描述一系列个人或组织,故而很少要求标准化访问技术。这类个案研究的典型做法是将受访者视为研究的合作者,告诉他们我们需要什么,请他们提供这些信息。研究者所要涉及的领域是特定的,但是随后可以自由提问,请受访者就这些主题进行进一步的解释,以便让访问员获得更多的信息。

与此相联系,有些社会研究的访问用于把握个体特征,这些访问在本质上并不能依据受访者的回答而达到目标。对于不能明确进行自我表述的受访者,描述性的回答只能用于分析,从而推导出有关结论(Mishler,1986)。例如,受访者可能被要求描述他们是如何应对压力环境,而研究者也许会从解决问题的角度,将这些处理方式分别编码为“建设性的”或者“破坏性的”。

在这些研究中,调查人员并不是将访问视为测量过程,而是一个信息收集过程。访问员一般会对问题进行解释和澄清,提出一些新问题,然后将受访者的回答进行汇总记录,而在标准化访问中这些都是被禁止的。进行这些访问时,为了尽量保持统一性,会使用数量相对比较少的访问员,而这些访问员完全了解研究目标。即使这样,研究者要将答案进行比较还是会出现问题,除非他们确实知道人们是在回答同样的问题,并在统一的框架之

内进行回答。因此,在非标准化的访问中加入某些标准化的问题也是有所裨益的。但是,问题的关键在于非标准化的访问要有合理的研究目标。

另一方面,有些看似针对标准化访问的批评,在我们看来其实是另有所指。例如,标准化访问通常高度依赖于封闭性问题或选项固定的问题。有些学者更倾向于用描述性回答来针对固定选项问题(如,Mishler,1986)。开放性回答或描述性回答确实能使受访者将其内心的想法更多地告诉研究者,但是,这类回答无法像封闭式回答那样能够提供进行量化分析的信息。而且,在将访问进行标准化管理时也可以设计一些问题并且采用叙述性的方式进行回答。无论以开放式还是封闭式的方式提问,答案的可解释性(interpretability of answers)才是普遍化的要求。我们认为这两类问题都有价值,如果研究目标是产生量化数据,那么,两者都应该进行标准化处理。

还有一个问题是,标准化对于收集事实信息和客观信息非常重要。关于主观状态、观点、感受或认识的回答,高度依赖于提问措辞的细节,研究者在这一点上几乎没有异议。但有些人宣称,在收集一些事实性信息时,比如,看医生的次数、住院次数,或者家庭成员的年龄,标准化问题并非如此重要。其实,即使最简单的问题都要求所有受访者对词汇和概念具有相同的理解。因此,当受访者被要求报告其年收入时,哪些收入来源应计算在内,哪些不应计算,哪段时间的收入应该报告,每个受访者都应该对这些有相同的理解,这点很重要。因此,对所有受访者使用同样的、定义完整的措辞,将保证研究者获得的资料是他们所需要的,并且可以从不同的受访者那里获得同样的信息。

然而,关于标准化的争论主要集中在当受访者对所询问的问题产生疑问时访问员应该如何做的问题之上。设计出来的问题对所有的受访者都同样明确,并且所指代的是同样的事物,这显然是不可能的。如果受访者对词汇的含义或问题的用意不能确定,访问员是否应该进行解释、说明或者对问题进行重新表述?还是应当由受访者自己决定这个问题的意思是什么?在标准化的坚持者与访问员创造性解决问题的鼓励者之间,这也许是争论最为尖锐的领域(Mishler,1986)。

最重要的一点在于，受访者对问题没有统一的理解，无论访问员做还是不做，都将产生测量上的问题。我们认为，如果问题设计有误，访问员就不能作出好的测量。处理这类有误问题的时机应当是在实地工作之前，而不是在调查开始之后的访问过程之中。

我们再来考虑前面的例子，即受访者被要求对其孩子就读学校进行评估的案例。我们完全可以假设研究者确实要求受访者从教育角度来进行评估，但却没有在问题中表述出来。因此，在我们的这个例子中，有些受访者可能会依据锻炼设施或者体育项目来进行评估。如果依据受访者的评论，他们可能误解了这个问题，或者受访者就问题的意思向访问员咨询时，有些人认为访问员可以澄清研究者的意图。我们认为，那将适得其反，并将导致测量更差，而不是更好。

1. 当访问员发现受访者不确定或者产生混淆时，如果他们可以自由改变或者修改问题，其结果就是，一些受访者被问到的是一个问题，而另外一些受访者被问到的是另外的问题，那么在进行分析时，也就没有人知道谁回答了什么样的问题。此外，如果受访者回答的问题不相同，答案就无法进行加总或者比较。
2. 除非一个研究者实施了所有的访问，或者访问员的团队人数非常少，我们认为，要求访问员超出问题提供界定或者解释以帮助受访者解决问题时，那么就很可能无法保持一致性。我们确信，事实上，如果要求访问员相机行事，访问员将产生众多与研究者目标并不一致的答案，因而不可能提高数据质量，使数据与研究者的目标保持一致。

社会科学研究有很多合适的方法用于访问，只有一部分研究要求使用标准化访问。然而，当目标在于产生定量数据时，需要的则是标准化的、非引导性的访问。

关于本书:使访问员相关误差最小化的途径

本书的主要目标是,描述访问员的特征和行为如何影响调查数据的质量,并说明研究者可以采取哪些可行的策略以减少调查测量中与访问员有关的误差。尽管越来越多的研究要求澄清数据采集方式对于访问员行为的影响,但我们在这里所讨论的这些原则可以应用于面对面的访问,也可以应用于电话访问。

关于减少或者控制访问员相关误差(Interviewer-related Error)的著作,要求对那些误差的性质进行了解。本书第2章将讨论与访问员有关的误差的性质及其估量。

所有调查机构都承认,在进行标准化访问时,访问员处理问—答过程的方式极其重要。提问、追问不完整的回答,以及记录答案都有一系列标准。这些程序都将在第3章进行描述。

实施访问的时候,问与答的过程本身在某种意义上是一环扣一环的。这是一个结构化的互动,在更大的场景中将访问员与受访者联系起来。此外,在大多数流行的调查中,访问员通过电话访问被调查者,或者登门拜访,从而建立起一种使这种访问得以发生的联系。事实上这种联系没有背景,也没有历史,访问员最多借助于事先的介绍信(advance letter),或者访问员的简短介绍和来意说明,随后就开始访问。

相对而言,作为访问的一个环节,交谈(speaking)往往被人们所忽视。但是,在过去20年中,Charles Cannel 和他的同事们进行了广泛的研究,讨论了访谈所得以发生的场景(context)对所获信息质量的重要性。他们集中讨论了访问员如何把握受访者的意图从而建立访问场景。第4章将讨论访问员—受访者的互动场景以及这一场景是如何影响调查数据质量的。

为了提高问卷调查的测量质量,研究者所采取的第一个步骤就是给访问员提供调查说明(script)。在早年的调查研究中,访问员仅仅被告知调查目标,并只接受一个如何从特定人群收集信息的培训就会被分派出去(Converse,1987)。很快人们就发现,不同的访问员会有各不相同的提问,这又会对答案产生重要的影响。

研究者注意到，如果将所要询问的问题写下来，让访问员提同样的问题，他们对所得结果的影响就会大为减少。尽管所有访问员应当询问同样的问题的观点已经在50年前就被普遍接受，但我们仍然认为，改进问题设计是控制与访问员有关的误差的最重要的方法。第5章介绍，现在如何设计调查问题能使访问员相关误差最小化。

论及如何最大程度地提高访问质量，每个人都建议使用优秀的访问员。一项更加深入的研究就是探讨访问员的特征，例如其宗教、种族、性别、年龄或教育程度是否会让回答产生误差。访问员的培训方式和督导方式也显然会影响他们的工作绩效及数据的采集质量。第6、7、8章总结了这些方面的现有知识，讨论访问员的挑选、培训和督导是如何影响访问质量和调查测量质量的。

在总结部分即第9章，我们将讨论如何使用标准化访问程序来实现访问员相关误差的最小化，并最大程度地提高以问卷调查为基础的测量质量。

第2章 什么是访问员相关误差

What Is Interviewer-Related Error?

对调查中的误差进行测量决非易事。从理论上讲，误差是指调查结论与研究者需要度量的“真实值”之间的差异。对于某些事实性问题，如受访者的年龄、看医生的次数，可以将调查结果与可靠的记录进行核对，但是在实践中，这种核对很少进行，甚至不具备可行性。而与主观现象有关的问题，如观点、感受或想法，甚至在理论上都无法对答案的准确性进行直接评估。

由于对误差不能直接进行评估，方法论学者便采用间接的方式。他们可以依据能够观察到的答案的差异大小，或者依据答案中不能合理地反映他们所要测量对象的平均差异，用来推断实际存在的测量误差。

测量者的个人特征和行为不应该影响测量结果，对此并没有争议。然而，依据访问员和答案之间联系的紧密性，我们知道，至少在测量过程中仍会存在很多误差。所谓访问员相关误差（Interviewer-related Error），我们指的是答案中的这样一些差异，这种差异与由谁进行访问有关。

我们同样知道，调查数据中的误差有很多原因与访问者个体特征无关，例如：

1. 提问的措辞（question wording）会影响回答；
2. 受访者个人特征中与测量内容无关的方面会影响回答；
3. 访问的情境会影响回答；
4. 特定的问题在整个问卷中所处的位置会影响回答；
5. 与受访者自填问卷相比，访问员的存在本身就会对回答产

生影响。

只要数据采集过程的任一环节对回答产生影响，就会形成误差。但是，本书并不直接涉及误差产生的其他来源。如果让访问员询问一系列问题，而在提问之时个体的访问员对答案没有影响，那么，在我们看来，就没有与访问员相关的误差。如果答案并没有测量出他们试图测量的东西，那么，解决方案在于更好的问题表达，或者是更好的操作程序，而不是改善访问过程。本书仅仅讨论回答的差异性，即来源于访问员个人特征的误差，以及访问员应如何做好工作。

标准化的重要性

依据我们对访问员相关误差的定义，毋庸赘言，标准化访问就是减少误差的方法。评估访问员是否标准化，就是看他们是否使回答受到了影响。如果评估表明，访问员影响了答案，那么，他们就没有标准化；如果没有影响答案，那么就已经标准化了。

但是，什么样的访问行为和流程才能最有效地保证数据采集过程的统一性呢？这里存在着许多经验方面的问题。显然，在真实的情景中，我们无法使访问员的行为在各方面都能够保持一致，我们的关键任务是找出访问员之间会对回答产生影响的差异，并尽量将其减至最小。为此，我们列出了一些旨在最大程度提高访问一致性的程序和技巧，检验这些方案的有效性，同时，仍须对访问员是否对回答产生影响进行评估。

对于读者而言，重要的是应当理解“访问员效应”(interviewer effect)对数据的影响，因为这种影响会妨碍研究者获得有效的结论。有两种方法可以观察访问员效应对数据的影响。这两种方法都是有效的，其中一种方法来源于心理测量学的测量路径，而另外一种方法来源于抽样统计学的测量路径。

心理测量学方法(psychometric approach)在本章导言中已经有所提及。关于社会科学中的测量质量有两个公认的标准，即信度和效度。信度是指一项测量所得结果的一致性程度；即如果反复测量同一特征或情况，所得到的结果应当相同。但如果由于访

问员的原因而使不同的受访者或不同的访问员之间出现差异,那么测量就会丧失一致性。所得到的回答不仅反映了测量对象的真实值与受访者回答本身的差异,也包含了访问员对回答所产生的影响。显然,访问员因素对答案的影响以各种方式扭曲了所要测量的东西,导致了不可靠的数据。表 2.1 对这一影响模式作出了简要的概括。

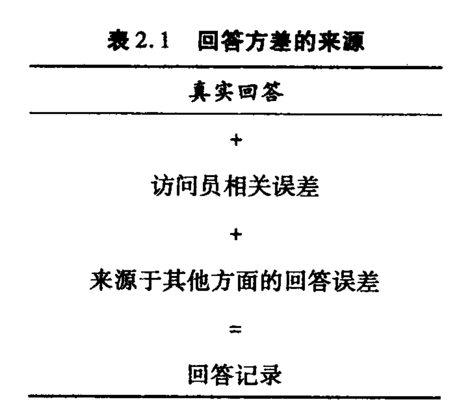


表 2.1 回答方差的来源

真实回答
+
访问员相关误差
+
来源于其他方面的回答误差
=
回答记录

社会科学中所使用的效度一词,指的是实际的测量与研究者的测量意图相吻合的程度。在理论上,依据调查结果与对同一事物的无误差测量之间的相关性或者一致性,可以对调查测量的效度进行评估。因此,某个调查结果与对同一事物或者相关事项的其他测量之间的相关程度,可以作为进行效度评估的标准(Cronbach and Meehl,1955)。

表 2.2 提供了公式,并总结了访问员相关误差与信度及效度之间的关系。一个简单的事实是,答案受访问员的影响越多,访问员相关误差越大,而调查结果与研究者要测量的事物之间的相关性就会降低;这是衡量效度的操作性指标,也是评判一项测量过程执行情况的关键标准。

表 2.2　信度对相关程度的影响

1. 观察方差(X) = 真实方差(X) + 误差方差(X)
   测量的方差包括两个部分:真实的方差以及由于不可靠的测量而引起的额外的方差。
2. 信度 = 1 - ($\sqrt{误差的方差}$/ $\sqrt{观察到的方差}$)
   计算测量信度的方法是,首先计算误差的方差占所观察到的方差的比率,然后用 1 减去这个值(该值小于 1.0),这样就给出了一个取值范围在 0 和 1.0 之间的信度系数。
3. "X"和"Y"之间的相关程度 = 真实相关程度 × $\sqrt{信度(x) \times 信度(y)}$
   由于每次测量的信度都取平方根,两次测量之间的相关程度会降低。
4. 效度(y) = "y"和标准"x"之间的相关程度。
   测量的效度取决于其与标准变量(criterion variable)之间的相关程度。

因此:访问员导致的误差降低了测量的信度,同时又限制了测量对有效性的反映。

抽样统计学家思考调查误差的角度多少有些不同。他们并没有集中于一个回答与某个效度标准之间的相关性,而是关注以调查为基础的估计量的精准度,例如均值和比例。他们考察两项内容,一是偏差(bias),即访问员所得答案的平均值与总体真实值间的差异程度;二是精准度(precision),是指以一次调查为基础的估计量如平均值,与对同一总体进行重复调查或者包括所有元素的无抽样调查(比如,数据来自于每一个体)所获得的估计量之间的相近程度。统计学家将这种对精准度的测量称作一次估计的"标准误(standard error)"。

估计偏差并不表明访问员缺乏标准化。使访问员更具统一性一般也不会减小平均偏差,而且在大部分内容中,本书并不关注调查估计中如何减少偏差。

标准化访问对抽样统计学家很重要,主要在于它会增加估计的标准误。访问员对回答的影响越大,特定样本的标准误也就越大。

为了对访问员进行研究,需要用统计量(statistic)来告诉我们访问员对数据的影响有多大。为此,Kish(1962)提出了组内相关系数(intraclass correlation coefficient),用以反映估计方差如平均

值,是如何与访问员相关的。因为,完全标准化的访问员应该与其所获回答没有关联性,组内相关系数为“零”意味着没有访问员效应;这一系数如果不等于零,则表示访问员与答案有关。

组内相关系数通常用 *rho* 表示。它用于测量访问员效应,与每个访问员所作的访问次数无关。因此,它可以用来比较不同组别和不同调查的访问员效应,也可以用于比较访问员对不同问题的影响程度。

广泛使用 *rho* 的一个实际限制因素是,只有在访问员所分配到子样本对总体样本具有代表性的条件下,才能进行计算,而且 *rho* 的计算公式十分复杂。

表2.3 给出的是 *rho* 计算公式中最为简单的一种情形。根据第一个等式,*rho* 表示一个估计量,如平均值的总方差中可以归因于访问员的那部分所占的比例。这一数据也告诉我们,在有潜在假定的情况下,如何依据方差的标准化分析结果计算 *rho* 的估计量。

表2.3 计算 *Rho* 及设计效应

A. *rho* 的概念是指估计量,如平均值的总体方差之中与访问员有关的方差所占比例。

$$RHO = \frac{\text{方差(访问员)}}{\text{抽样方差} + \text{方差(访问员)}}$$

B. 使用将访问者作为随机效应变量的方差分析结果,可以依据模拟均方(model mean square)、误差均方(error mean square)以及每个访问员的平均访问数量来计算 *rho*。

$$RHO = \frac{\text{Model MS} - \text{Error MS}}{\text{Model MS} + (n-1)(\text{Error MS})}$$

C. 由于访问员对回答结果具有影响,*DEFT*(“设计效应”的均方根)是标准误的膨胀(inflation of the standard error),大于从总体中抽样的标准误。其值取决于 *rho* 的大小以及访问员所做访问的平均数量。

$$\underset{(DEFT)}{\sqrt{\text{Design Effect}}} = \sqrt{1 + (n-1)(Rho)}$$

* 只有在如下条件满足时,这一设计才适用:(1)访问员所访问的子样本,必须可以代表整个样本;(2)每个访问员访问的数量应该接近于相等。关于如何计算 *rho* 及其前提假定的充分讨论,读者可以参考 Kish(1962),Stokes 和 Yeh(1988)以及 Groves 和 Magilavy(1980)。

一般而言,*rho* 的显著值一般在 0.01 到 0.02 的区间内(虽然有时会高达 0.10)。这一数值表示,这些估计量的方差中有百分之一或百分之二是由访问员引起的。表面看来,这些数值并不构成大问题。但关键在于,这一效应是每个访问员的访问数量的乘数。一个访问员访问的数量越多,他们对数据的影响也就越突出,越有显著性(参见表 2.4)。

表 2.4 不同 *Rho* 值和访问员平均访问数量的条件下,计算访问员效应*对于均值标准误的估计值(*DEFTS*)的乘数效应

每个访问员的平均访问数量	组内相关系数(*Rho*)				
	.005	.01	.015	.02	.03
11	1.002	1.05	1.07	1.10	1.14
21	1.05	1.10	1.14	1.18	1.26
31	1.07	1.14	1.20	1.26	1.38
51	1.12	1.22	1.32	1.41	1.58
81	1.18	1.34	1.48	1.61	1.84
101	1.22	1.41	1.58	1.73	2.00

注释:* 由样本规模和样本设计而计算出的标准误估计值,应当以表中所载访问员效应之系数相乘而有所膨胀。

设计效应的统计量是以数字来表示这一效应,它通常以平方根的形式出现,表示为 *deft*。表 2.3 中最下面就是 *deft* 的计算公式,它是 *rho* 值与每位访问员所作访问数量的平均值的函数。表 2.4 列出了不同 *rho* 值和访问员平均访问数量条件下的 *deft* 计算值。例如,如果 *rho* 为 0.015,访问员平均访问数量约为 50,则标准误的估计量需乘以 1.32;这表示,在同一调查中,内部相关系数为 0.015 的标准误的估计量,比内部相关系数为 0.0 的标准误的估计量要大 32%。

要在 *deft* 值与 *rho* 值、抽样统计学家对访问员效应的测量与心理测量学家(psychometrician)对信度和效度的测量之间建立简

单的数学联系,很不容易。准确的计算还取决于调查设计中方差的其他来源。但是,调查测量的效度与 *deft*,即估计量的标准误有(反向)相关关系。关键在于,从两个角度来看,访问员对回答的影响越大,与访问员有关的方差就越大,数据在进行精确估计以及就相互关系作出有效结论方面的价值也就越小。

我们应该指出,标准化访问程序还有一个比降低访问员相关误差的重要性更为关键的价值,即可重复性。社会科学的一个原则是研究发现应当可以被其他研究者所重复发现。如果访问员在某种程度上以其自身的方式提出问题,自行向受访者表述问题并处理受访者的问题,研究者就无法告知其他研究者这些数据是如何收集的。如果研究者无法重复某个具体的发现,人们就很难判断这究竟是一个结论不稳定的简单问题,还是由于数据采集的差异性对结果产生了影响的问题。

有证据明确表明,缺乏标准化的访问程序将会导致显著的、可以识别的访问员相关误差。但是,只有很少的研究者可以设计研究,从而估计访问员效应对其数据的影响。但是,即使没有对访问员效应进行测量,最大程度地要求访问员以统一的、标准化的方式采集数据,仍应当成为高质量的研究工作的一部分,这样才能为其他人重复其研究发现提供基础。

检测访问员相关误差

发现访问员相关误差相当不易,需要有特殊的测量方法才能将其检测出来。如果没有加以特别的注意,人们很容易推断访问员对数据没有影响。下述三种不同的方法可用于识别访问员相关误差,但每种方法各有其优势与不足:

1. *对访问员进行直接观察*。通过安排观察员在一旁观察或通过录音记录,我们可以了解访问员是否按照标准化的方式进行操作。通过这种观察,人们可以检测访问员是否按照事先写好的问题进行提问,他们的后续提问是否恰当,以及对答案的记录是否准确。本书所提及的很多发现都来自于对访问现场的观察或录音记录进行的研究(如 Hyman, 1954; Cannell, Fowler 和 Marquis, 1968; Fowler 和

Mangione，1986）。

但是，对访问员的行为进行观察，仍无法提供信息帮助人们判断访问员给数据带来了多大程度上的误差。虽然对答案的记录误差可以得到准确评估，但是，由于提问失误或追问欠妥而产生的影响却不能直接评估。例如，错读问题，经常但却不一定会引起误差。因此，这些观察研究只能表明访问员做的那些事情在我们看来会影响答案，但是不能提供直接证据说明他们在多大程度上影响了所收集的数据。

2. *寻找访问员与回答之间的联系*。这是识别访问员效应的第二种方法。如果访问员的做法符合标准化要求，我们就不能将答案与谁做的访问联系起来。一些研究使用了前文所述的内部相关系数来简单论证是否能够根据访问员本身去预见其所得到的回答（如，Kish，1962；Fowler 和 Mangione，1986；Groves 和 Kahn，1979；Groves 和 Magilavy，1980；Stokes，1986a；1986b；Stokes 和 Yeh，1988）。另外一些研究则试图找寻访问员个人特征与其所获回答之间的联系（如，Schuman 和 Converse，1971）。

这些研究所需的特定程序在一般调查中并不常用。具体地讲，如果受访者是随意地或为了方便而进行分配的，那么访问员的个人特质就不可能与其所分配到的样本毫无关系。为了就访问员对数据的影响进行有意义的分析，每个访问员所分配到的子样本对整个样本必须具有代表性（至少是可定义的一部分）。由于这在设计中一般无法做到，因此多数研究者无法通过对数据结果的检验来判断访问员是否对回答产生了影响。

3. *检验调查回答的有效性*。在可以通过有效渠道获得"真实"的答案时，可以将此作为评估调查误差的第三种方法。有些研究结果可以用来与公共记录进行对比，如投票登记、酒后驾车记录（Locander，等，1976），或者与一些医疗卫生事件进行对比，如住院记录和看医生的次数（如 Cannell，Marquis 和 Laurent，1977a）。但一般情形下，人们进行调查是为了得到目前所缺乏的信息。人们在进行有效性核对

之时,只有一小部分信息是可以进行对比的,而且这些样本对总体不具有代表性;因此,尽管这些研究被证明对我们的方法论知识有所贡献,但是此类研究数量是相对有限的,而且同样存在着局限性。

本书所报告的结论来源于以上所有三种类型的研究。尽管这些研究的数量相对偏少,但是一旦得以进行,它们就会提供大量的证据可证明访问员是调查误差的一个重要潜在诱因。例如,在对访问员程序进行考察的研究文献中,Cannell(1968)发现,进行医疗卫生事务访问调查的访问员所使用的追问,有35%可以归类为诱导性追问,有可能对回答产生影响。Groves 和 Kahn(1979)就如何对访问员效应进行计算出版了一个摘要,他们发现,这些研究所使用的问题中,大概有1/3左右,其与访问员相关的内部相关系数在0.015或更高。如表2.3所示,如果每位访问员平均访问50次,这种内部相关性的标准误差将增加32%。显然,具有相对比较高的内部相关系数的问题,其对标准误的影响更大。

从另外一个角度出发的研究中,Sanders(1962)指出,一项关于健康卫生状况的调查的数据中,有超过50%的方差可以归因于访问员。

作为一项运用记录核对的研究,Cannell 等人(1977a)发现,在另一项健康调查中,访问员承担访问数量的多少与其所报告的住院率之间的相关系数达到0.72。

结 论

依据上述检测访问员相关误差的方法,已有大量的证据表明,访问员是调查中一个重要的误差来源。研究组织和研究者很容易忽视这类访问员相关误差,因为如果不采用特殊的方法进行识别,它们并不明显。毋庸置疑,访问员缺乏标准化是大批调查数据达不到其应有质量的一个原因。此外,与某些减少调查误差的策略如扩大样本容量相比,降低访问员相关误差的成本通常非常低廉。接下来,我们将讨论如何最大限度地减少调查中的访问员相关误差。

第3章

标准化访问技术

Standardized Interviewing Techniques

要执行一项高质量的标准化调查，让所有访问员都统一访问方式，的确非常困难。但是，让访问员按照标准化的方式来处理问—答过程，却可以简单表述为如下几个程序：

1. 按照字面表述方式准确无误地读出问题。
2. 如果回答者对于最初问题的回答不完整，或者不恰切，应该按照非诱导的方式进行追问，让其进行澄清并阐述细节，也就是说，以一种对回答内容没有影响的方式进行追问。
3. 对答案的记录应该没有加入访问员的主观判断，答案应当而且只能反映回答者说了什么。
4. 访问员应该以中立的、无偏见的方式处理答案的实质内容；关于访问所涵盖的主题，访问员不应提供任何可能暗含特定的价值观或者偏好的个人信息；在涉及答案的特定内容时，访问员不应该对受访者做出任何反馈，无论是积极的反馈还是消极的反馈。

但是，在以标准化方式实际操作时，却有两个方面的主要障碍：

1. 不恰当的调查工具。如果事先设计的问卷不易按照标准化的方式进行控制，那么，就不可能执行标准化的程序。这个主题我们将在第5章中进一步说明。
2. 受访者不理解他们要做的事情是什么。调查中的测量过

程是一项团队作业,需要所有的参与者都按照指定的方式扮演角色。访问员难以按照正确的方式完成其工作,一个常见原因在于他们不能够,或不知道如何指导受访者从而使访问过程运转起来。这个主题将在本章稍后进行说明。

此外还有两个因素加剧了访问员进行标准化访问的困难,当他们在处理问卷或者应对访问者时存在麻烦的情况下,尤其如此。这两个因素是:

1. 访问员一般要求准确的答案,因此当标准化的目标与准确性看起来存在冲突时,他们就难以被标准化。
2. 访问员喜欢以友好的方式对待受访者,当他们感到培训时的要求与受访者所期待的关切之间发生矛盾时,他们往往就会难以实现标准化的访谈。

实际上,标准化的目标并不一定必然与获得准确的数据以及对受访者作出应对发生矛盾。本章最后一部分将讨论如何在不忽视其他重要目标的情况下实现标准化访问。表 3.1 概括了标准化访问的技术及其障碍。

表 3.1　标准化访问的技术及其障碍

标准化访问的技术:

1. 按照书面表述读出问题。
2. 对不完整的答案进行非诱导性的追问。
3. 记录答案不进行擅自取舍。
4. 在涉及答案实质内容时不进行个人化的评价。

标准化访问的障碍:

1. 不恰当的调查工具。
2. 受访者不知道如何扮演其角色。

访问员难以标准化的原因:

1. 准确性目标似乎与标准化目标相矛盾。
2. 亲和性目标看起来与标准化目标相冲突。

按照字面表述方式读出问题

实际上，我们所分析的每一份访问员手册都将"依据书面表述的方式读出问题"作为高质量访问技术的第一项基本原则。表面上看，这似乎是一个很容易理解并得到遵守的规则。因此，发现访问员往往没有按照书面表述的方式提问，这也许会让人们感到惊讶。

在对访问员—受访者互动过程进行编码的四项研究中，访问员实际将书面表述的提问方式作出改变的比率在 20% ~ 40%（Bradburn & Sudman, 1979；Cannell, Fowler & Marquis, 1968；Fowler & Mangione, 1986；Cannell and Oksenberg, 1988）。而且，必须指出的是，这些研究都是由一些对方法论严格性比较强调的机构所完成的，因而关于访问员实际改变提问方式的比率，所得到的这些数字可能是比较保守的。

为什么访问员会改变字面表述？当然，在一些案例中，问题的设计者应当承担主要责任。如果问题很难念出，访问员就有可能改变表述方式。他们也会改变提问方式以强调某些部分，如果他们认为这将有助于受访者理解问题或者理解研究者要询问什么内容。但是，这些解释只能说明访问员失误的一部分原因。在我们看来，让访问员改变字面提问方式的主要动力在于他们力图使互动过程更加具有交流性，更加自然。他们那样做的一种方式是将他们个人的感受加入问题。随着访问员的职业发展，这一行为将有可能继续发展并且日益严重；除非访问员在没有按照字面方式准确读出问题时，督导立即加以控制并进行反馈。Bradburn 和 Sudman（1979）发现，相对于新入行的访问员，经验丰富的访问员更有可能比较随意地进行提问。

当然，在大多数调查项目中，对问题表述方式做出改变的情形看起来都非常少，而大多数访问员也会说他们基本上按照书面表述的方式读出了问题。从测量的角度来看，关键问题在于改变提问方式是否会导致测量质量出现差异化。尽管我们还不能确定，但是我们确实知道，改变书面的提问方式能够对答案产生非常大的影响。

我们考察这个问题的一种方法是,对于那些最有可能被访问员误解的问题,考察其依据内部相关关系进行测量时是否最有可能产生访问员效应。但在这方面,我们的发现是否定性的。当访问员被要求就一个很难读出的问题进行提问时,他们所做出的改变并没有对数据产生重要的访问员效应。

另外一种方法是看那些比较随意提问的访问员是否使数据产生显著的偏差。我们的评估方法是对20位访问员的访问进行录音,然后进行编码。因为每位访问员的样本都是总样本的一个概率性的子样本,考察基于一个访问员所访问的子样本的估计与依据总样本进行的估计是否有差异,这是有意义的。我们简单地计算了每个访问员样本中我们所认为具有潜在偏差的样本,也就是说,与总样本的均值相比具有偏差的样本。这一数字与访问员在标准化访问技术方面的表现得分存在着相关性。

我们发现,对不同技能的评分,包括读出问题、追问和进行记录,均存在着内部相关性,因此我们不能将好的提问与好的追问所作的贡献区分开来。由于只涉及了20位访问员,粗相关系数也不具有统计显著性。但是,与所有这些标准化行为的相关关系的方向,包括与读出问题这一行为的相关关系的方向,则验证了我们的预期。这就是说,那些在访问技能上面,包括提问技术上面得分比较高的访问员,更有可能获得偏误比较少的数据。

回答该问题的第三种方法是对问题的字面表述进行研究。Schuman 和 Presser(1981)介绍了许多实验,这些实验有意识地对问题的字面表述方式进行细微改动,然后对回答结果进行比较。总地来说,他们发现,某些时候字面表述方式的细微改变会导致答案的分布发生巨大的变化;而在另外的一些案例中,提问方式的显著改变对人们所给答案的影响却非常小。

问题 A:美国是否应该禁止在公共场合谈论共产主义?

问题 B:美国是否应该不允许在公共场合谈论共产主义?

人们可能会指出,"禁止(forbidden)"和"不允许(not allowed)"是相同的概念,但是对回答者来说这两个概念却不相同。例如,在对两种提问形式的样本进行比较时,近50%的被调查者回答美国应该"不允许"在公共场合谈论共产主义,但是只有20%

的被调查者回答美国应该“禁止”人们在公共场合谈论。这一发现表明，如果访问员在提问时用“禁止”来替代“不允许”，将会对数据产生巨大的影响，从而产生明显的访问员相关误差。对访问员来说，这一改变可能看起来无伤大雅而且有利于沟通，使提问变得更好，但是却给我们提供了一个很好的案例，说明为什么要告诉访问员应该按照书面表述的方式读出问题。

访问员往往被教导说要认真念出所有的选择项。但访问员往往容易增加或者删除某一个选择项，这就是说“或者您在那个议题上没有意见”。Schuman 和 Presser 为我们提供了一个与选择项有关的很好的案例。几个问卷样本均要求人们回答是否赞成或者反对一个假想的《1978 年农业贸易法案》(*Agricultural Trade Act of* 1978)。其中一个样本明确提供了这样一个选项，“或者您在那个议题上没有意见”；而另外一个样本没有提供这样的选项，而只是询问人们是否喜欢或者反对那个法案。结果表明，有 69% 的被调查者自觉回答他们不知道答案，而当调查提供了“无意见”这个选项时候，有 90% 的被调查者选择“无意见”。

是否包括一个“无意见”的选项，会对答案的分布产生重要影响。没有读出一个选项看起来是一个很小的改变，让访问员更容易提问，但是这一行为却产生了访问员相关误差。

另一方面，Schuman 和 Presser 也报告了其他一些实验结果，访问员对表述进行了很大的改变却对答案没有产生什么影响。例如，用“堕胎”(abortion)一词来代替“终止怀孕(end a pregnancy)”一词，对人们给出的答案并没有什么影响。

总的来说，让访问员完全按照书面表述的方式进行提问，在理论上似乎很容易理解。如果访问员没有按照他们的书面表述方式提问，研究者便不能确切知道他们究竟提了什么问题。但是，一般性地教导访问员不要改变字面的表述方式，还是不够的。访问员往往会改变提问方式，除非有一些重要的措施阻止他们这么做。此外，随着时间的延长，他们更有可能更多地这样做，除非这个趋势受到监控。访问员改变提问方式的动机可能是无害的(innocuous)，甚至是有建设性的，意在使问题更清楚，使他们与受访者的互动更自然，简单来说是要帮助研究者改进工作。有些时候，这些改变是无伤大雅的。但是，有些时候他们却造成了巨大

的差异。总的来说,研究者在设计问题时候要尽量让其可以按照字面表述方式诵读出来。确实,如果要进行严格的测量,这就是一项必须做到的工作。

对不完整的答案进行追问

在理想的情境下,研究者设计出高质量的问题,访问员则按照书面表述的方式进行提问,而受访者提供完整的答案,从而达到设问的目的。当然,这样的情况并没有经常出现。如果最初的提问没有得到令人满意的答案,访问员必须采取一些措施继续这一过程从而达到设想的目标。这就意味着访问员的行为无法做到完全程序化,因为每种情境中的要解决的问题都不一样。但是,我们的目标是让访问员按照统一的方式处理问题,不会因为访问员和受访者的不同而对回答的内容产生影响。这一行为,包括其间的几个步骤,可以被称为"非诱导性追问"(non-directive probing)。

实际上,访问员必须解决两种类型的问题。首先,他们有时候必须向受访者澄清问题的意义;其次,他们必须激励受访者补充、澄清甚至以其他方式修改他们的原始答案,从而完成并符合设问目标。

从原则上讲,对问题进行澄清应该是比较简单的。最为经常出现的情况是,如果受访者不能立即明白问题,那只不过是未能把握这一问题的某些方面。访问员的反应应该是重新完整地念一遍问题,强调受访者在开始时忽略了的词汇或者某些部分。尽管第二遍读问题的时候,间或会忽略指导语(introductory phrase)或者指导手册,但访问员必须确保读出整个问题,这样,受访者在准备回答时才会在事实上获得了与其他受访者同样的刺激。

当然,在某些情况下,症结并不在于受访者没有听到提问,而是提问中的某些词汇或者概念的意义对受访者来说比较模糊。但即使如此,我们仍认为访问员也不能以此为依据进行随意取舍。如果这个词汇在问题中有定义,访问员可以重新读出这个定义;如果该词汇在问题中没有定义,那么受访者必须依据他们自己的最佳理解来给出回答。

我们知道,研究者在撰写培训手册时会就某些可能模糊的词汇进行界定,这样,访问员在遇到受访者问询时,就可以使用这些界定。我们毫不置疑地认为,那只是一些并不完全的解决方案。首先,访问员不会在访问过程之中打开培训手册查找研究者所提供的准确解释。他们可能会依据记忆对定义进行重新解释,但是这样的解释在不同的访问员和受访者之间很不一致,从而形成了非标准化的提问。更糟糕的情况是,只有那些询问了访问员或表现出明显疑惑的受访者,才会获得这些特别的界定,而其他受访者(包括那些实际上同样疑惑的受访者)却不会获知这一定义。

当问题涉及一个显然很关键的词汇,它没有很好的定义却有可能产生多种含义,也绝对不能让受访者与访问员进行商讨。当受访者被问及去看精神分析师是否应该算做看医生时,"你怎么考虑都可以"可能是一个严格的回答。但是,解决这个问题的出路是在数据采集开始之前,而不是在访问过程之中;解决这个问题的人是研究者,而不是访问者。一旦数据采集工作已经启动,访问员就应该以标准化的方式对他们所拿到的问题进行统一的表述,即使有些概念界定并不清晰,也同样作出统一表述,这样才能进行最好的测量。

追问的另外一个目标是获得符合设问目标的答案。访问员的行为在某种程度上依赖于受访者的回答行为。受访者的回答行为通常可分为三种类型:从问卷所提供的一系列选项中选择一个选项(涉及封闭性问题);提供数据型答案;以受访者自己的话进行回答(一般涉及开放性问题)。

对封闭式问题进行追问

如果一个问题要求受访者从一组答案中选择一个,而受访者却没有这么做,那么访问员的工作就是向受访者解释,回答问题的方法就是从所列选项中选择一个(称为指导受访者),并再一次复述所列选项。

访问员在面临这一情况时往往有两种类型的错误。第一种情况是访问员接受一个并不正好符合某个选项的答案,然后自己将这个答案归为某个选项。简单说,访问员在代替受访者选择一个答案。

人们比较容易理解这样的情况:

访问员:您如何评价您的学校——很好,比较好,一般,或者比较差?

受访者:这个附近的学校不是很好。

在这点上,人们可以理解访问员为什么应该核对选项并且进行追问。当然,问题在于有些访问员可能选择"一般(fair)",而另外一些访问员可能选择"比较差(poor)"。如果访问员采纳受访者的措辞并进行调整,进而产生一个答案,那么,不同的访问员之间的不统一的情况将非常普遍。而当受访者选择一个答案,当然不能确保完全没有误差,但仍可以使答案仅仅取决于受访者,而与访问员无关。这正是标准化要达到的目标。

在对封闭式问题进行追问的时候,访问员的另外一种错误是在选择项需要复述的时候,他们没有重复所有的选择项。在上述案例中,当受访者说"不是很好"的时候,如果访问员追问:"哦,您是在说一般还是比较差?"这是可以理解的,但仍是一个会影响答案的不好的追问。我们能够证明,关于"比较好,一般,比较差"这个量表(scale)的答案分布,与关于"很好,比较好,一般,比较差"这个量表的答案分布是非常不同的。受访者在将其自身的选择进行归类时,不仅会对选项数量以及某个选项在量表中的位置进行反馈,也会对措辞进行反馈。将一组答案只截取一部分,将会使受访者获得的信息极不相同,这将会对答案产生影响。

对数据型答案进行追问

当一个问题的答案要求填写数据而非进行单项选择时,访问员所面临的最普遍问题是精确性。受访者往往用一个数值区间或者概数进行回答,此时访问员会让受访者回答得更为精确一些。

一个不合适的追问行为是诱导性追问。进行诱导性追问时,某个答案得到选择的可能性会高于其他答案。访问员进行诱导性追问的方式有很多,但最容易的方法是要求访问员回答"是"或者"否",从而对某个答案进行确认。这样的追问被认为具有诱导性,其根本原因在于它可能提示了某个特定的答案。如果这样提问,受访者更有可能回答"是"而不是"否"。因此,任何可以用

"是"或"否"进行回答的追问都具有诱导性。此外,任何列举或者提及某些可能的答案而排除另外一些答案的追问也具有诱导性,因为它提高了所提及的答案被选中的可能性。

要求提供数据型答案的问题往往要求访问员追问更多的特定细节,而这样就变成了诱导性追问:

提问:在过去 7 个晚上,您睡眠时间少于 8 个小时的有几次?

回答:我一般都睡 8 个小时。

诱导性追问 1:哦,在过去 7 个晚上,您的答案是 0?

诱导性追问 2:哦,在过去 7 个晚上,最合适的答案是 0,1,或者 2?

非诱导性追问 1:在过去 7 个晚上,您睡眠时间少于 8 个小时的有几次?

非诱导性追问 2:哦,在过去一周,最合适的答案是超过 2 次,还是少于 2 次?

跟进性的非诱导性追问 2:(如果答案是少于 2)哦,在过去 7 天中,最合适的答案是 0,1,或者 2?

前面两个追问的问题显然在于它们提示了一个答案。很显然,第一个追问最差,因为这个追问可以用"是"与"否"来回答;第二个追问稍微好一些,但是访问员已经限定了受访者选择的空间,在答案不同于 0,1 或者 2 时,受访者也不会主动采取行为选择其他的答案。

在这个案例中,最好的追问是对问题进行复述,因为受访者提供答案的真正问题在于没有回答所提问题。作为最标准化的方法,复述问题具有独特的优势,因为访问员没有创新或者创造问题。

从非诱导性的角度来看,非诱导性追问 2 是一个可以接受的回应方法,虽然它尚不够标准化。在这个案例中,访问员使用了一种可以成为"零定位(zeroing in)"的技术:访问员依据一般情况进行合理的猜测,然后提出一个并没有提示性的问题,答案可能会在截断点(cutpoint)的一端或者另外一端。如果从最初的问题中获得了一个答案,跟进的问题就用于进一步限制回答的区间。

某些时候,“零定位”是能让受访者回答得更为精确的唯一方法,或者是最有效的方法。但是,我们自己的偏好是让受访者而非访问员或者研究者估计精确答案所在的区间,然后让受访者依据他或者她自己的情况进一步明确化。简而言之,一般来说,我们的策略是对受访者进行指导,并坚持书面表述的问题。

对开放性问题进行追问

对访问员来说,最难的任务是对开放性问题进行追问。访问员对于已经获得的答案必须进行如下三个判断:这是否回答了问题? 答案是否明晰? 答案是否完整?

通过查阅各种访问员手册,我们发现,较之于要求依据文字表述方式进行提问的规定,不同的机构在对访问员如何追问进行指导时,差异会更加明显。有些机构似乎认可,甚至鼓励访问员通过各种交谈方式来让受访者阐明或细化先前的回答。我们的偏好则是让访问员用少数几种方式进行追问。

事实上,我们培训访问员时,除了要求重复问题,他们只需追问三个问题:

1. 您想表达什么意思?
2. 请您谈详细一些。
3. 还有补充吗?

这三个追问容易记忆,且不具有诱导性。它们不会给访问员以任何机会进行发挥,这样就不会使访问因受访者或访问员的不同而产生差异。我们认为,如果组织者鼓励发挥性的追问,那么所获答案将会千奇百怪。而由于缺乏标准化,交谈的兴趣也会发生变化。此外,假如访问员的自由追问确实是非诱导性的,他们就可能属于上面所提到的三个问题或者其中的一个问题。

访问员的任务就是从中决定哪种追问是合适的,这需要分析受访者提供的答案。如下四种追问,包括重复问题,分别适用于受访者回答不完整的四种情形:

1. 答案没有回答问题,但回答了其他问题;访问员应将问题重复一遍。

2. 回答中包含有不明晰的概念或措辞,表义模糊;访问员应该追问"您想表达什么意思?"

3. 回答不够详细或具体;访问员应该追问:"您能够说得更详细些吗?"

4. 访问员已经给出了一个完美贴切的答案,但是受访者在答案中有可能再补充一些;访问员应追问:"您还有补充吗?"

以下是这几种追问方法的运用实例。

提问:在您看来,居住在这个社区最好的方面(best things)是什么?

点评:这是访问员和受访者要处理的最棘手的问题之一,因为所谓"方面"根本缺乏具体说明。能够提及社区的哪些特点,提得多详细,都取决于访问员和受访者的取舍。

回答1:我们以前居住的那个社区流动性很高,人们不在乎同社区的联系。

点评:这个回答的毛病在于没有回答问题。虽然依据对以前社区的描述可以对现在社区的特点进行推断,但问题是要对当前的小区进行描述。

追问:复述问题。

回答2:人。

点评:这似乎是对问题的回答,但是没人能弄清楚其含意。我们需要详细说明。

追问:请您谈详细一些。

回答3:这里的人是好邻居。

点评:有人会认为这就是完整的回答,但实际上很难说。如前面所述,卷面的问题在于没有给出线索使访问员和受访者都知道什么样的回答才能满足研究者的需要。这是否是一个充分具体的答案则取决于问题的目标设定和编码程序。但是,什么才是"好邻居",不同的受访者会有不同的看法,而迄今为止还没有提供任何相关信息。一个好的访问员可能

会去追问,从而更多理解受访者的想法。

追问:您所说的好邻居是什么意思呢?

回答4:他们自己管自己的事情,让你独自呆着。你不用要求让自己变得更合群,也不用去担心别人在想什么。

点评:这一出人意料的回答体现出了追问的价值。依据开始的回答,有人会预期对这个社区所谓"好邻居"的描述是如何很热心、友善、乐于助人。我们现在知道这位受访者所提到的人和好邻居是指什么了。不过,这个问题是让受访者说出这个社区的几个比较好的"方面"。

追问:好的,我记下了。还有补充的吗?

点评:如果一个问题没有限制受访者要表达观点的数量,访问员应该继续追问"还有要补充的吗",直到受访者说"没有了"。

对回答不知道者进行追问

如果受访者回答说"我不知道",那么对于访问员而言,这要求有特殊的追问。"我不知道"对于知识型的问题是一个合理的回答。但它也可能意味着:

A. 这是受访者的回应风格,是他或者她思考答案的开始。

B. 受访者之前从未思考过这个问题,但是如果他或者她仔细考虑,就可能得到答案。

C. 受访者知道回答,但不确定答案是否足够具体和精确,是否合乎研究者的要求。

如果受访者说"不知道",访问员的首要任务就是尝试诊断难回答的原因。

A. 如果"不知道"是对信息类问题的准确而且认真的回答,那么访问员应该记下答案并转入下一个问题。

B. 如果这是受访者延缓回答的方式,那么访问员应该给受访者一些时间让他或她进行思考。访问员也可以复述问题以帮助受访者思考。

C. 如果受访者以前没有考虑过这个问题，那么访问员应该鼓励其思考，并强调他或她在这个主题上有独特的资格提供信息，然后复述问题。

D. 如果受访者不能确保回答的质量和准确性，那么访问员帮助其进行重新确认：回答没有对错之分，设计这些问题只是用来了解每个人自己的观点和看法；受访者自己最好的估计，比没有提供任何信息要好。然后访问员复述问题。

追问错误的类型

追问无疑是访问员最难掌握的技巧。访问员在追问过程中会犯两类错误：诱导性追问以及对于需要追问的答案没有进行追问。

最初，研究者主要担心访问员会进行诱导性追问，从而使答案按照他们所期望的方式呈现出来。例如，他们认为，共和党的访问员将进行诱导式追问，从而提高回答支持共和党观点的受访者数量。但是，访问员受过良好的训练后，这种情况便不会发生（例如，Hyman，et al.，1954）。多数情况是，访问员进行诱导性追问在于他们认为他们已经知道受访者想要提供的答案，但却不知道如何让受访者明确表达出来。

*诱导性追问*是调节访问员—受访者互动的一种策略。访问员发现，如果他们对一个回答追问了数次仍然没有让受访者提供符合问题目标的答案，就会有一种紧张。当受访者已经说了很多，访问员和受访者都非常确信访问员已经知道答案，对访问员而言最简便的事情就是说："我认为您要回答 X，对吗？"

访问员的另一类错误是没有追问应该被追问的答案，或者在决定哪些回答应当进行追问而哪些回答不进行追问时前后并不一致。人们发现有三种情形特别容易导致访问员的追问不一致。

第一，Hyman 等人（1954）发现，访问员的预期会影响到追问行为。具体来说，如果访问员所得到的回答与自己所预期的回答一致，访问员就会接受答案而不进行进一步追问，而这种预期是基于对受访者的了解以及受访者对先前问题的回答。但是，如果受访者提供的答案与访问员预期不一致，他们就可能进行追问以

确保他们获得正确答案。这是尽责的访问员行为所导致的对答案处理的不一致。

在询问可能有多个答案的问题时,访问员所获得的答案数量存在着差别。在可能有多个答案的情况下,访问员须自主决定什么时候追问"还有补充吗"。而面对同样的问题,有些访问员获得了更多的答案,因为他们连续追问,这样也对数据产生了影响。

第三,访问员在处理"我不知道"或类似回答的方式上存在着差别。在受访者最初不知道如何回答时,有些访问员或者更为努力,或者更为成功地取得了观点或者答案。

追问是提问—回答过程的一个组成部分,无法完全做到标准化。如果受访者在第一次提问时没有提供完整的回答,访问员可以作出追问或不追问的决定,回答也无法完全程序化。只要访问员有机会自行判断,不同的访问员和受访者之间就可能会产生不一致性,这就可能导致访问员相关误差。

追问是不是调查测量误差的重要来源呢? 毫无疑问。在 Hyman 的早期研究中,追问差异是访问员行为导致误差的主要环节。正如上文所提及的两个问题,处理"不知道"答案的方式,以及受访者所提到的答案的数量,一直是与访问员效应有关的两个问题(Groves 和 Magilavy, 1980)。在我们自己的研究中,追问开放性问题的质量与访问员的答案偏差之间的相关关系接近于显著性水平($p=0.07$)。此外,我们在第 5 章中还要讨论,与访问员效应联系最紧密的问题可能正是那些要求进行追问的问题。

正如我们在第 5 章中将会指出的,我们认为,试图将追问因素导致的访问员相关误差降到最低,最有效的方法就是提高问题质量。访问员必须进行的追问越少,产生误差的机会也就越小。此外,我们强烈主张将访问员的自由追问降到最低,访问员采用创造性的方式来刺激受访者进行回答的机会越多,出现不统一和产生误差的可能性也就越大。而且,在访问过程中访问员除了要考虑追问以外,还有很多事情要做。好的问题不仅会降低追问的可能性,而且在需要进行追问时,还可以降低追问所产生的不一致性。

记录答案

访问员的工作是记录受访者给出的答案。标准化记录的关键在于不让访问员进行判断、归纳，不让其影响记录内容。开放性问题和封闭性问题之间，要求提供事实性信息的问题和旨在了解观点和感受方面信息的问题之间，标准化记录的规则都有所不同（见表3.2）。

表3.2 不同类型问题的答案记录原则

开放性问题、了解事实

记录所有与提问目标相关的信息。

开放性问题、了解观点

逐字逐句记录回答，不作任何说明和归纳。

封闭性问题，了解事实

核对受访者所选答案。

如果受访者不确定哪个选项合适，则将问题作为开放性问题处理，记录所有相关信息。

如何处理答案留待编码操作之时再作决定。

封闭性问题，了解观点

核对受访者所选答案。

进行追问，直到受访者作出选择。

不要选择答案选项，除非受访者进行选择。

对于封闭性问题，访问员的关键任务在于让受访者选择一个答案，然后进行核对或者记录。除了笔误之外，记录出错的情况唯一可能就是访问员自己认为某个答案是受访者选择的，而事实却并非如此。

关于观点或者态度的开放性回答的记录规则同样是清楚而简单的：访问员应当逐字记录答案，也就是说，访问员应当记录受访者说的原话，不进行概括或者删减。文献表明，不同的访问员所作的概括和解释会出现差异（如 Hyman et al.，1954）。防止访问员对回答产生影响的方法就是减少其在记录答案时加入的主观判断。

如果问题询问的是事实性信息,不论是开放性还是封闭性问题,规则都稍有不同。对于主观性回答的编码在很大程度上取决于回答的表述方式以及受访者表达或未表达出的某些特定观点,而事实性问题一般是询问某些具体信息。受访者的措辞并非关键,访问员只需记下受访者所提供的与问题有关的信息,而不必记录原话。此外,如果问题包括了一系列选项,而受访者又没有进行选择,访问员最好记录受访者所提供的相关信息,而不一定非要受访者选择一个选项。

询问事实性问题的一个重要特征就是,我们难以预料访问所可能遇到的所有特殊情形。此外,受访者可能会对某些概念和定义进行询问或者表示疑问,而这些概念在设问之时并未阐明但却对受访者给出确切回答至关重要;访问员在这时提供定义或者帮助并不是好的方法,因为这个定义或者帮助并非可以统一提供给其他受访者。在需要处理一些没有预计到的特殊情况时,访问员采取一些特殊方式进行弥补也是不容许的。对于事实性问题,最好的方法是对影响答案的关键概念进行各种解释,然后收集回答问题所需要的所有信息。这样研究者就可以制订统一的编码规则去处理所有受访者的情况。

前文我们举例讨论了看医生的问题。有一名受访者想知道,只是找护士进行注射是否算看医生。如果访问员或者受访者都无法依据问题而进行判断,最理想的解决方式是把两种情形下的回答都记录下来。因此,访问员应该记录:"看过 4 次医生,另外去了 10 次,仅仅是找护士进行注射。"

有关记录的另外一个普遍问题,我们可以用询问受访者接受了多少正规教育的例子加以说明。高中之后的受教育经历并不都能严格对应为正规的教育序列。假设问题提供了如下选项:高中毕业以下,高中毕业,上过大学(some college),大学毕业。如果受访者自述在高中之后的教育还包括有艺术、音乐、护理、汽车机械,以及其他内容,那么,让访问员(以及受访者)来决定这些是否属于大学教育经历是不可行的。访问员要做的就是记录下受访者所说的教育模式,而让编码者来处理这些问题,从而保证所有访问员和受访者之间的一致性。

因此,对事实性问题而言,标准化测量的障碍就是访问员自

作主张。逐字记录并没有那么重要,但是,将所有信息都记录下来,从而使决策规则能得到仔细考虑并统一应用,这是标准化测量的关键。

我们对访问员的研究表明,他们在记录封闭性问题的回答时误差相对较少,而对于开放性的观点类问题进行逐字记录时,则质量差异很大。这需要访问员花费很多努力去仔细记录,但如果督导不坚持这样做,访问员一般不会这么做。然而,这又非常重要,因为访问员的概括和解释是非标准化的。一如已经提及的那样,Hyman 等(1954)发现访问员的预期会影响他们的记录内容;访问员倾向于记录那些与他们对受访者的认识比较吻合的答案,他们的解释和概括会使答案相互矛盾而内容单薄。

实现人际中立

所有的访问员手册都鼓励人际中立(be interpersonally neutral),将其作为标准化过程的一个部分。有时候人们很难准确说明这是什么意思,但是我们认为,如果将其加以限定,应该包括如下行为:

1. 访问员不应主动向受访者提供自身有关生活状况、观点或价值观的个人信息,特别不能提及那些可能与访问主题相关的任何特征。绝大多数调查机构都会要求将此类交谈减至最少,这至少有三个方面原因:第一,访问的目标是建立一种专业性而非私人性的关系,其首要任务是采集数据,而主动提供个人信息可能会有损这一目标。第二,尽管访问员不会被认为具有可以观察到的人口学特征,但谈论个人处境和观点只会强化不同的访问之间的差异。第三,关于个人观点和背景的信息可能会直接影响回答;最严重的情况是,受访者可能会猜测哪个答案最符合访问员的偏好和价值观。
2. 在访问互动过程中,访问员应当注意,受访者提供的回答不能包含任何对回答内容的评估或判断;访问员的目标就是获得准确、完整的答案。很自然,受访者会关心访问员

如何看待这些答案。访问员应该注意,不能在随意的人际交流中进行任何评价。

在标准化访问的这一方面,受过一些训练的访问员都能做得很好,而未受训练的访问员则表现得不尽如人意。依据我们的研究,在几乎未受训练的访问员中,有超过三分之一的人被认为需要在处理人际交流问题上进行提高;而那些受过较多训练的访问员中,超过85%的人在处理人际交流这方面被评价为“满意”或比较好(参见第7章)。

这些数据与Hyman的研究发现相一致。他关注于访问员是否会用自己的个人观点与受访者进行沟通;他发现受访者通常认为,他们对访问员的态度一无所知;而当他们确实认为自己知道访问员的想法时,这些想法往往并不准确。显然,如果受访者认为他们了解访问员的观点,他们往往就认为访问员会同意他们的观点,而不再考虑访问员的实际观点。

在我们对访问过程进行观察时,进行很夸张的价值性评论的情形非常少见。在培训中,我们会讲一个访问员如何问受访者喝多少酒的故事。当受访者说他每天喝六杯时,访问员惊呼“我的天,这太可怕了!”,这种事件在一个认真规划的调查中一般不会发生。

难以把握的是反馈过程中的某些细微环节。例如,受访者说:“我已经有一年多没有去看医生了。我想我真的很幸运。”这时,访问员可能会说:“那不是很好么。”

这有什么不对吗?难道我们没有普遍的人道主义,不希望所有的人身体健康吗?当然,这是对的。但是,在回答下一个问题时,如果受访者必须报告一些并不那么健康的情况,他可能会认为访问员将低估他,或者感到不高兴,或者不知道怎么关心人,那么这就不是我们所希望的了。在这种短暂的关系中,受访者会寻找各种线索来保证不出差错,而那些认为受访者身体很好或者保养得很好的、细微的、无足轻重的欣喜表达,可能对数据产生影响。

这一点有多重要?很难给出确切的评估。但是,Marquis, Cannell和Lqurent(1972)的研究显示,对受访者行为的细微反应

会显著影响到他们所报告的就医次数。此外，根据我们的研究，在录音访谈中，访问员对受访者的不当反馈与回答所产生的偏差之间有显著的相关性。

研究表明，访问员一般能够成功地避免明显夸张性的评价，绝大部分认真运作的调查都不会出现这个问题。但是，访问员和受访者之间更为细微的交流却很可能会在相当大程度上对回答产生影响。

指导受访者

通过研究我们相信，访问员在作标准化访问的时候，最重要的步骤之一就是对受访者进行指导。调查问卷设计不好固然是困扰访问员的一个主要原因，但是，当调查开始后，访问以什么样的方式进行，以及在什么样的规则下进行操作，才真正成为令人困扰的问题。访问员的一个应对措施是改变标准化规则，使自己在受访者面前更有亲和力，让受访者感觉更加舒服。而我们认为，更好的解决方式是向受访者解释，说明为什么要以标准化方式进行，而不是用非标准化的方式进行。我们确信，如果访问员坚持这么做，测量的质量将会有显著提高。

指导受访者有两种基本方式。一是在访问开始时提供简要介绍；二是在访问过程中出现问题的时候，就标准化访问的一些细节进行解释。从实践上讲，两者结合是最好的。

我们强烈建议访问员在访问开始之前朗读如下一段话：

> 因为很多人从来没有接受过这样的访问，我朗读一段话，把访问方式向您介绍一下。我会按照问卷内容向您提出一系列问题，这样，所有的受访者都会被问到完全相同的问题。您需要回答两类问题。有些问题，您将用自己的语言来进行回答。对于这些问题，我会将您的答案逐字逐句记录下来。其他一些问题，您需要从已经列出的答案中选择一个最合适的答案。在访问过程中，您如果对回答要求有任何不明白的地方，请向我提问。

这个简短介绍完成了几件重要的工作，可以有利于访问的进

行。首先,它介绍了这样一个事实:这是一次特殊的交流,有一系列特殊规则。这与受访者此前所经历过的其他互动形式,包括其他类型的访问都不相同;同时,它还告知受访者,在访问进行的过程中,访问员还会对访问规则进行进一步的阐述和解释。

其次,它提前告诉受访者访问员将会怎样做,这也有助于访问的进行。此外,如果访问员告诉受访者他将会严格遵照原文朗读问题,这也就增加了访问员不这么去做的难度,由此也就增加了访问的标准化。

访问开始后,如果受访者不能很好地适应角色,我们认为访问员应当停止问—答程序,向受访者解释规则,并强调为什么这些规则很重要。以下是最常见的几个难点:

难点:访问员还没有提问,受访者就开始提供部分答案,甚至完整的答案。访问员如果继续念这个问题会感到难为情,因为这样就会让人觉得他没有听取受访者早先的回答。

访问员:"下面这个问题您已经做了一些回答。但是,访问规则要求我请您逐一回答问题,这样我们可以将您的回答与其他人进行比较。另外,我们还发现,虽然某些问题之前已经回答过了,但是对特定的问题答案会有所不同。所以,我逐字逐句念一下这个问题,请您进行回答,这样保证我们不会出差错。"

难点:受访者认为问题中的某个词语意义模糊,或者没有很好地界定,而问题表述本身没有就受访者所关心的方面提供充分的解释。

访问员:"我明白您是对提问表示质疑。尽管这些问题都是经过严格测试的,但仍然会有个别问题在某些人看来不甚明了,或者说不是对每个人都适用。但是,在调查过程中,我们要求人们依据问题的字面表述方式提供最合适的答案。这样我们可以将您的回答和其他人的进行比较。如果我们针对每个受访者都修正提问,我们就不能对回答进行分析。我再读一遍问题,请您按照它的表述,给出一个您认为是最佳的、最准确的回答。"

难点：在封闭型问题中，受访者不想从所提供的选项中选择某一项。

访问员："对于这种问题，是依据人们所选择的选项来对回答进行分析。我请您在其中选择一个，这样我们可以将您的回答与其他人提供的答案进行比较。我们知道，有时候没有任何选项能确实反映您的情况，但是，其他人可能也会遇到同样的问题。最重要的是，我们要让对每位受访者的问—答过程都保持一致，这样我们才可以比较出不同回答者之间的异同。"

难点：受访者所给出的回答不够具体，而仅仅是估计或猜测。

访问员："我们希望您能够提供最好的估计。虽然这个估计可能不够精确，但是没人会比您更能作出这项评估。请您尽力。"

难点：受访者语速太快，给访问员的逐字记录造成了困难。

访问员："为了保证记录准确，我必须逐字记下您的回答。如果我自己总结，可能就会产生偏差。对于这个问题，请您适当放慢一点语速。我可能会让您重复某些部分，以免出错或是遗漏。"

难点：家庭其他成员想要帮助受访者回答观点性问题。

访问员："对于事实性问题，例如您看过几次医生，住过几次院，如有需要您可以向别人询问，因为我们需要最准确的信息。但是，如果问的是观点或是感受，那么除了您之外没有人能给出回答。同样，这关系到每个人所接受的访问是否一致。如果我们询问人们的感受和想法，受访者本人独立回答是最合适的。虽然很多人都对他人很了解，但是我们认为其他人无法确切地告诉我们某人的想法或者感受。因此，为了保持调查的一致性，我们必须确保这种问题是由受访者本人来回答。"

难点：在访问中，受访者征求访问员的意见。

访问员:"访问结束后,我可以谈论任何您关心的事情,但在访问结束之前则不可以。因为我们发现,在某些情况下,如果访问员在访问中表达自己的观点和想法,我们就会影响到受访者的回答。我们所设计的访问过程,唯一对答案有影响的就是您的情况和您所说的话。"

我们还可以继续列举很多难点及其解决之道。但这些具体情况并不重要,重要的是过程。如果一名访问员试图进行标准化的访问,而受访者却不知道如何扮演角色,结果就是一个糟糕的互动过程,访问员就会被迫在标准化和积极回应受访者这样两种情况中频繁进行选择。的确,为了对受访者进行有效的指导,访问员必须得到足够的信息,他们必须了解访问规则以及遵循这些规则的合理原因。大多数访问员在自己接受培训过程中都能获得这些解释,关键在于必须确保他们在访问过程中也能正确使用这些规则。

结 论

以标准化方式实施调查确实是一件困难的任务。从技巧的角度来看,以一个前后统一且非诱导性的方式进行追问无疑是最困难的一项。这是最缺乏恰当标准的一项,也是最容易导致访问员相关误差的环节。记录开放性问题的回答同样很难做好,也是误差的一个来源,虽然标准化调查中已经减少了对开放性问题的使用,因而与过去相比,这个方面出现的问题少了一些。

访问员不喜欢严格遵照原文进行提问,他们喜欢自己对问题进行修改。作为误差的一个来源,提问失误可能没有追问失误那么重要。已经出版的研究成果表明,访问员自行改动问题的比例并没有严重到损害标准化的程度。此外,在那些比较随意操作的调查中,访问员改动问题的比例要高于研究文献所报告的情况,后果也更严重。尽管我们很难确切了解这一情况的严重性,但在这里必须重申的是,进行标准化测量的基本前提就是,我们要知道受访者究竟被问到了什么问题。

关于保持中立,访问员似乎都能理解并且欣然接受。本书的

观点可能是有偏差的，因为研究访问员的那些人所接触的都是有组织的调查团队，这些组织都有明确的操作标准并对此十分重视。访问者只要有一个小小的松懈就很容易使交流行为带有个人的判断和评价，因此我们不得不怀疑，在实际操作中，这个问题可能要比数据所显示的更为严重。

最后，我们改进标准化访问的一个最重要工作就是强调指导受访者的重要性。通过我们的观察，标准化中断的一个主要原因就是受访者不能进行合作，或者不能理解访问过程。如同我们跟人下棋，对方觉得自己只是搬动棋子这么简单；双方都明白下棋大致是怎么回事，但如果对具体的规则认识不清就会造成一个十分尴尬的局面。调查访问也与此相似。受访者对于访问都有一个大致的了解，但是他们对于一个特定的标准化访问应当怎样进行却并不了解；让他们知道不同的规则，并在访问进行的过程中因势利导地向他们解释这些规则，将使访问的质量大为提高。

第4章 为标准化访问创造情境

Establishing the Context for Standardized Interviews

成功实施标准化访问的一个前提是建立某种关系，这样，在测量的过程中受访者会愿意、而访问员能够承担自己的角色。

关于访问的大多数思考和写作都聚集于问—答过程的质量，以及如何保证质量。然而，Charles Cannell 及其同事的重要贡献在于，将注意力集中于问—答过程得以进行的情境问题，并且指出，访问员在执行访问这种测量试验过程中有着关键的作用。毫无疑问，问—答过程的处理方式对于测量质量非常重要，这在第3章已经讨论过。但是，同样毋庸置疑的是，访问员如何建立访问互动，如何处理与受访者的关系，从而使问—答过程得以进行，也对访问数据有着关键性的影响。因此，访问员肩负着重要责任。这是访问员工作中要求极其苛严的一个方面，也是访问员很难保持一致性的一个方面。

在本章中，我们将讨论建立问—答互动的三个重要方面：

1. 解释访问的目的或原因。受访者接受访问必定有其原因，他们接受访问也必定会导致一定的结果。在本章的第一节，我们讨论的是受访者为何会接受访问，不同的原因和动机对数据结果又有何影响。
2. 在访问员和受访者之间建立相互协调的关系。一方面，这种关系要足够积极，这样受访者才会乐意接收访问并保持配合；另一方面，双方都必须明白这种访问关系强调和重视的是准确且完整的回答。在本章的第二节我们将介绍如何建立起这样的关系及其这种关系对测量质量所产生

的影响。

3. 介绍访问目的和将要适用的标准。具体来讲，受访者需要怎样配合，任务的完成应当有多高程度的严谨，这些标准都是由访问员掌控和制订的。研究证明，这方面的互动是对测量结果影响最大的一个方面。本章的最后一节将着重讨论这一问题。

受访者接受调查的原因

大多数访问都属于一种"瞬时关系（instant relationship）"，也就是说，受访者事先对这一调查访问几乎没有任何了解或背景知识，访问员要在很短的时间内对访问工作进行简单的介绍，解释访问目的，与受访者建立起某种关系，对受访者进行指导。

不管是入户调查还是电话访问，在开始接触访问员之前，受访者可能已经收到了一封介绍信（advance letter），向他们介绍这个研究课题的目的及其资助者。近几年，越来越多（很可能占大多数）的访问都是通过随机数字拨号系统来完成的。这样一来就无法事先发出介绍信；在有人接听电话回答问题之前，研究人员无法得知究竟哪户人家会被抽中进行访问。

设想一下下面这个场景：

访问员：您好，请问是(617)956-1150 吗？

受访者：是的。

访问员：请问这是住宅电话还是公司电话？

受访者：是住宅电话。为什么问这个？

访问员：您好，我叫 Mary Smith。我在州立大学的调查研究中心工作。我们正在进行一项有关人们对艾滋病关注程度的研究，这项调查是由联邦政府资助的。我们要在全州范围内抽取一部分成年人进行调查。为了确认您的住所中哪位成员应当接收访问，我需要了解和您同住的人中有几位年龄在 18 岁以上？（注释：很多调查对于受访者的选取有严格的规定，因此第一、二个问题通常会与住所成员有关。）

受访者：我一个人住。

访问员:那么您就是我需要访问的人。如果可以,我们现在就开始访问?

在上述这样一个简单的开场白之后,访问员会对调查项目以及受访者需要做哪些事情进行进一步的介绍,随后开始向受访者提问。

毋庸赘言,受访者随后会被问到什么问题以及这些问题有何用途,是各不相同的。调查的赞助者可以是商品的制造商和销售商、政治候选人、报纸、州政府或联邦政府以及各种基金会。调查的内容几乎无所不包。从受访者正在使用的牙膏牌子到他的投票倾向,乃至他是不是有过加剧艾滋病风险的行为,任何问题都有可能被问到。访问的长度亦有可能从几分钟到两个多小时不等。

对受访者的要求也有差异。有些问题询问的是印象或感想,这就要求受访者稍作思考;而有些调查则需要采集有关整个家庭成员当年健康状况的详细信息,这就需要受访者查看病历或是询问其他家庭成员,以保证数据的准确性;还有的问题需要访问员从所给选项中进行选择,有的则需要受访者用自己的话语来回答,想到什么就说什么。这些调查所起的作用可能是帮助某个牙膏制造商提高销量,可能是帮助某个候选人当选,可能是为医疗保险政策的制定提供基本信息,也可能是评估某个健康保险项目的成本。不管用于何种目的,不管受访者被问到什么问题,研究人员都希望访问过程能提供良好的测量结果,所得到的回答都能精确有效地反映被调查人群的情况。

不同的访问对受访者的要求也不一样,衡量的标准之一是看访问所需耗费的时间和精力;另一个方面则是看受访者是不是需要回答有关个人隐私或是有可能产生后顾之忧的问题,比如使用违禁药物的情况或是一些不雅的个人习惯。受访者是否能够很好地配合,在一定程度上取决于其参与的动力和访问的氛围。

Kelman 受访者分类法

研究受访者对于访问的态度,一个有用方法就是 Kelman (1953)对影响关系的研究。显然,问—答双方的互动就是影响情

境的一个例子，这个关系由访问员建立，而受访者根据访问员的指示进行配合。

Kelman 区分了三种影响关系，其动力各不相同：

服从型关系(compliance)是指根据受访者的表现灵活运用赏罚来施加影响的关系。比如在一次访问中受访者不会说“不”，不会无视访问员，从而得以避免不愉快场面的出现，那么这就是一个很好的顺从关系；如果一名受访者的首要想法是同访问员搞好关系，他想要得到表扬或享受访问的乐趣，那么这也算是一种顺从关系。采用这种方式鼓励受访者参与访问的一个主要问题在于，评判受访者表现优劣的主要标准即回答的准确性，一般不易测得。一个看似表现很好的受访者很有可能根本是在做无用功。因此，奖励的方式不是总能得到令研究人员满意的表现。

认同型关系(identification)是 Kelman 所指称的第二种影响关系。人们认可自己在某种关系中所扮演的角色，并按照这个角色的行为预期来采取行动。一名受访者所要扮演的角色可能会有几种之多。在社会科学调查中，一名好市民为公益事业尽一份力的想法可能是最突出的一个角色预期；另一个想法可能没有这么“高尚”，但起到的作用是一样的，那就是受访者想要表现出和善的态度，当别人提出合理要求时尽量配合并给予帮助。这种关系对获得良好的访问结果能起何种作用，在很大程度上取决于受访者对扮演这一角色有多少后顾之忧。

内化型关系(internalization)则是 Kelman 所说的第三种影响关系。这一影响关系得以产生的原因是，人们会将特定的行为与自己的价值观联系在一起。在访问中，有的受访者会把参与访问看成是实现个人目标的一个途径，比如提供社会服务，或是了解一些有关重要社会问题的情况。与认同型关系一样，内化型关系能否带来好的访问结果，取决于受访者是不是认为提供完整而准确的回答是实现其目标的最好方式。

受访者对访问的定向

要评估 Kelman 的分类法是否有用，我们就需要了解受访者对访问的态度。这里我们借用两项调查的结果。在这两次调查中，

受访者都接受了一个30分钟的健康状况调查;之后的一到两天内,我们又对其进行了回访,询问他们对于先前那个健康状况调查的感想。这两次健康调查在程序上是十分相似的:都事先寄出信函告知受访者研究的目的,随后由访问员进行入户调查。但是这两项研究的主要目的和赞助方有所不同。密歇根大学在1960年代所进行的研究追踪的是美国健康访问调查(Cannell, Fowler和Marquis, 1965b)。这次普查是由人口普查局的访问员完成的,依照合同帮助收集公共健康服务方面的全国性数据。波士顿地区的这项研究(Fowler和Mangione, 1986)针对的是受雇于马萨诸塞州立大学调查研究中心的访问员,他们得到联邦政府的资助,负责为当地健康规划采集数据。这两项研究向受访者提出的问题是各不相同的,但是比较下来,两项调查所呈现的问题大致相同。

原理1 受访者对一项调查的了解程度差别很大,很多人对于自己所参与的调查由谁出资、目的何在几乎一无所知。最惊人的表现是密歇根大学关于美国健康访问调查的研究所发现的结果。如前所述,受访者在访问前收到过一封介绍信,在访问中有访问员进行标准的指导,访问后又收到一张介绍出资方的感谢信。尽管这样,在第二天接受回访调查的时候,近半的受访者宣称不知道之前那次访问的目的何在,同样有近半的人不知道访问员服务于哪家机构。事实上,尽管人口普查局要算是这个国家中最知名的调查机构了,但是只有约十分之一的受访者能正确地说出它的名字(见表4.1)。

表4.1 不同受教育程度受访者对于健康访问调查信息的了解

	0—8年级学校	1—3年高中	4年高中	大学1年及以上
对研究目标的了解程度				
清楚	27%	45%	41%	67%
不清楚	14	13	15	7
不知道	59	42	44	26
	100%	100%	100%	100%
由谁聘请的访问员				

续表

	0—8 年级学校	1—3 年高中	4 年高中	大学 1 年及以上
普查局	2%	9%	14%	18%
联邦卫生局	13	11	22	42
卫生署	15	15	15	13
政府	8	2	12	0
其他	3	3	4	6
不知道	59	60	33	21
	100%	100%	100%	100%
N	(129)	(89)	(123)	(67)

资料来源:依据 Fowler(1966),略有调整。

受访者的回答与其受教育程度有很大关系。这项研究完成于 1960 年代,鉴于从 1960 年代至今美国受教育程度整体有了很大的提高,因而现在的受访者对所受调查的了解程度在平均水平上应该要高于 1960 年代的那次调查。上过大学的人当中,有三分之二对研究目的能够有清楚的认识,有 60% 的人知道受访者受雇于人口普查局或某个联邦卫生机构。然而,大多数调查研究对于研究本身及其资助方的介绍都不够重视,因此几乎可以肯定地说,在一般的调查中,很多人对调查本身不会有太多的了解。

原理 2　访问在受访者生活中并非重要事件。例证同样来自密歇根大学对健康访问调查项目的研究。在回访中他们会询问受访者,在调查访问结束以后是否想起过这回事,是否和亲戚朋友们谈起过这次访问。大约 40% 的人说访问员走后就再没想起过这件事;同样有 40% 的人说自己没跟任何人谈起过这次访问。

在波士顿地区所进行的调查则采用了更为直接的测量方式,来询问受访者认为所参与的访问的重要性。44% 的人认为"非常重要",47% 的人认为"比较重要",9% 的人认为"不重要"。

与受访者对调查的了解程度一样,他们参与调查的原因也各不相同。最直接的信息还是来自密歇根大学关于健康调查访问项目的研究。他们向受访者展示了一幅图画,画上有一名受访者正在门口与访问员谈话,请受访者谈论此时他们觉得画中的受访

者对于接受访问有哪些积极和消极的想法(参见表4.2)。

表4.2

人们为什么会对一项调查作出积极反应

积极因素	提及的积极因素所占百分比
提供帮助或者服务	35%
喜欢与访问员进行交谈	35%
喜欢问题的内容	5%
希望获得个人的好处	11%
正好在休息	14%
	100% = N(314)

人们为什么会对一项调查作出消极反应

消极因素	提及的消极因素所占百分比
很忙—耽误太多的时间	37%
调查不值得	14%
对调查目的不了解	29%
对问卷内容表示质疑	20%
	100% = N(796)

资料来源:根据 Cannell et al. (1965b),略有修改。

积极反应中最突出的两项是想要提供帮助和希望与访问员交流。消极反应中最普遍的一条(把表4.2的中间两行合起来看)是受访者觉得接受访问对自己而言没什么好处;其次最为普遍的原因,是受访者很忙或者还有其他事情要做;对于问题本身的担忧(无论是私人问题还是不好回答的问题),则最少被提及。但对于健康以外的其他访问内容而言,情况可能有所不同。

把这些数据结合起来看,可以得出一个重要的结论,即人们参与调查的原因以及他们对访问的态度是不一致的。我们知道,改善受访者行为作息的质量,是永无止境的;但我们也知道,有两种尝试是不会成功的。

一个是密歇根大学(Cannell, Fowler 和 Marquis, 1965c)提出的。他们请了一位专业画家设计了提前发放的介绍材料,使其内容丰富,更能吸引受访者的眼球。这些材料随后被寄往受访者手

中，他们最后还会接受访问，谈谈自己在收到这些材料后的反应。只有大约一半的受访者说“认真读了”而不是“扫一眼”；而大约三分之一的人根本不记得收到过这份材料。新式传单对阅读率的提高没有什么明显的作用。

与此类似的是药品的说明书。药品说明书的目的同样是尽量使患者在一个相对很短的时间内对该药品有一个必要的了解。研究显示，在经过一个标准的指示过程后，患者对药品的了解程度仍存在着很大差异（见 Meisel 和 Roth，1983）。

虽然可能有更有效的方法向受访者提供信息，但我们的结论仍然是：研究人员必须要假定受访者对访问的了解程度是参差不齐的。这样的话，受访者对访问的态度也是不同的。

回到先前我们讨论的影响模型，关键的问题在于，受访者对于访问的所想、所感、所知是否会影响到所得数据的质量。虽然相关的数据并不明确，但是现有证据倾向于认为这种影响并不大。

密歇根大学的案例还进一步研究了受访者所说的积极和消极因素与访问质量之间的关系（Cannell，Fowler 和 Marquis，1968）。证据表明，如果在受访者时间紧、不方便的时候进行访问，比如就餐时间，那么访问的效果就会更差。但是，没有证据表明，那些表示有兴趣与访问员进行交流的受访者与没有这方面兴趣的人相比，访问质量有什么不同；同样也没有证据表明，那些表示想要为公共服务尽一份力的人与没有这样说的人相比，访问质量要更好。高中以上学历的受访者在更多地了解研究目的及资助方以后，会提供更多的健康信息，但这一趋势不是很明显。然而尽管，这一相关性并非十分明显，但在那些没有完成高中学业的被调查者中，根本不存在这种相关性（Fowler，1966）。

虽然没有直接证据，但我们仍然相信，服从是最不能激励受访者参与访问的一个原因。很简单，访问员很难根据受访者回答的质量来进行赏罚。如果受访者参与访问是为了达成某些直接的个人目的，比如有偿接受访问或是仅仅出于礼貌不愿拒绝，而不是为了提供准确完整的回答，那么结果可能事倍功半，因为这种态度可能带来精确的回答，同样也会产生失真的数据。但是，就我们的目前所知而言，受访者是出于个人动机（内化型）参与访

问还是为了与访问员搞好关系(认同型)而参与访问,对数据结果的质量并没有多大差别。我们的观点是,不管受访者参与调查的原因何在,关键是要使他们明白,他们的任务就是好好回答问题。本章的最后一部分将会讨论如何实现这一目标。

访问员—受访者关系

上面的讨论强调了在许多受访者眼中问答双方关系的重要性。由于受访者对访问的了解程度参差不齐,大多数受访者愿不愿接受访问的关键可能就在于他们与访问员的关系。而且,与访问员的这一关系将会对受访者在测量过程中的表现产生重大的影响。

我们已经说过,在大多数调查中,问答双方的互动关系是在短短几分钟之内建立的。而在电话访问中,受访者连与访问员见面的机会都没有,只能通过声音来建构对访问员的印象,通过对方的话语来了解自己所要扮演的角色。在这种情况下,受访者几乎肯定会通过自己熟悉的某种角色关系的类比来定位这一情境。此时,他们所选取的参照角色就显得至关重要。

这里很关键的一点就是,问答双方理想的角色关系包含有几个不同的、甚至相互矛盾的方面。一方面,访问员应当要营造一个信任和尊敬的气氛,要让受访者相信这次访问是为了一个正当合理的原因而进行的,访问员应当向受访者保证不会带来任何风险和伤害。此外,我们希望访问员是主导交流的一方,因此他们最好看起来既有一定威信,又显得平易近人。这样的话,访问员最好表现得像一名教师、医生、律师或社会工作者。

另一方面,访问员经常需要向受访者提出一些涉及隐私或让人难以启齿的问题,有时甚至带有风险或威胁。这种情况下最合适的角色模型是什么呢? 有人说是朋友,因为人们往往会和朋友交流那些难以向陌生人开口的事情。但是在短短两分钟内,访问员几乎不可能同受访者建立起无话不谈的朋友关系。另一种可能合适的关系是邻居。但是,有时候受访者可能更愿意向完全陌生的专业人士提供个人信息,而不是自己抬头不见低头见的人。

然而,另一个需要注意的方面是,访问需要在一个能够进行

自由交流的气氛中进行。像医生、教师这样一些具有威信的人物形象可能会让受访者觉得自己必须给出严密正确的答案，这样反而会使他们觉得拘谨，不利于问题的回答。

总的来说，我们需要的是一个随和而不失专业的关系，访问员能得到尊敬和信任，但是这种专业性是聆听性的，而不是进行评判。

这一描述可能让我们想到了心理医生或咨询师，这并不奇怪；术语乃至临床专家，如 Carl Rogers 所说的技术，在标准化访问调查的术语和执行中均开始扮演重要角色。当然，人们肯定还会质疑，在某些随机拨号之后进行电话访问之时，大多数访问员是否能够成功地建立这种虚拟的临床关系。

同样，我们所掌握的只有两项研究是针对访问员所建立的问—答关系及其对数据质量的影响。此外，它们的研究对象都是入户调查。我们必须据此推测，与这些发现相比，电话访问或以其他方式进行的访问有何异同之处。

对于"受访者眼中的访问员是什么形象"这一问题，密歇根大学针对健康访问调查的研究给出了一个回答。表 4.3 显示的是受访者认为访问员最像和最不像哪类人。研究者可以有信心地指出，访问员所塑造的主导形象是专业性的；社会工作者、护士和教师是最常见的三个回答。不过，也有不同的回答。19% 的人认为访问员最像一个邻居或好友，而几乎相同数量的人认为访问员最不像邻居或好友。而与一些预测恰恰相反的是，大多数受访者都不认为访问员与上门推销的销售员有什么相似性。

表 4.3 受访者对于访问员的形象的报告

访问员是：	最像	其次像	最不像
专业性角色			
护士	7%	13%	3%
社会工作者	46	16	2
女医生	2	5	6
教师	3	13	5
女律师	2	2	21
办事员与销售人员			

续表

访问员是:	最像	其次像	最不像
秘书或办公室职员	14	15	4
商店里的女售货员	1	4	12
上门推销员	3	6	17
非专业性角色			
邻居	12	9	6
社区中心的志愿者	1	6	7
亲密朋友	7	7	15
不确定	2	4	2
N =412	100%	100%	100%

资料来源:根据 Cannell(1965b,表 8.7,第 77 页),略有调整。

如果说访问员最常展示的形象是一名专业人士,那么对于他们应当如何行事,至今尚无定论。事实上,密歇根大学的调查显示,受访者几乎被平均分成两类,一类认为访问员应当表现得"例行公事",另一类则希望访问员表现得像"登门拜访"(见表 4.4)。对这一问题的问答与受教育程度有关;受教育程度低的受访者更加倾向于访问员是"登门拜访"。

表 4.4 对第 37 题的回答:"有些人更希望访问员具有专业性——专注于工作——但是,其他人希望访问员有点像登门拜访者。你最喜欢哪一种角色?"

访问员应该如何行事?	
只强调与工作有关的问题	36%
比较多地强调与工作有关的问题	3
两者同等强调	12
比较多地强调登门拜访者角色	5
只强调登门拜访者的角色	39
其他	0
不确定	5
N =412	100%

资料来源:根据 Cannell(1965b,表 8.8,第 79 页),略有调整。

尽管在访问风格的喜好上存在着明显的差异，但是一个压倒性的结论是，受访者都表示很喜欢对他们进行访问的访问员。例如在密歇根大学的研究中，88% 的受访者表示自己遇到的访问员在表现友善和公事公办的分寸上把握得刚刚好。波士顿地区研究对访问员的专业性和友善性给出了一个从 1 到 10 的量表，而受访者打出的平均分都在 9 分以上；有超过 90% 的人认为访问员做得“出色”（58%）或“很好”（33%）。虽然受访者明显有不愿意批评访问员的倾向，但这两项研究的数据都表明，绝大部分受访者对访问员的评价还是很高的。

如果受访者对访问员的总体印象很好，我们却仍然要问，是不是有某些行为模式相比而言更能够带来较好的回答效果呢？答案是“可能有”。

在密歇根大学的研究中，访问员在以下几个方面存在着差异：一是当受访者向与访问无关的方向转移话题时，访问员是否能及时引导；二是访问员是否显得幽默；三是访问员是否表现得与受访者建立起了私人关系，比如，在访问结束后留下来交谈一两分钟。如果说这些做法能够使访问气氛更加随和、更能调动受访者的积极性的话，那么的确有一些证据表明，在对高中以下学历受访者的健康调查中，这些做法有助于回答质量的提高；但是，对于高中以上学历的受访者则没有这一效果（Fowler，1966）。在波士顿地区进行的研究显示，那些使受访者觉得更为友善的访问员所得到的回答比较不容易产生偏差。但是，Weiss（1968）的一项研究（第 6 章将要详细讨论）表明，如果访问员对受访者表现得过度亲善，数据结果就会产生较大的偏差。Hensen（1973）将访问风格进行融合处理，让访问员时而表现得亲和友善，时而更具专业性；结果表明这种方式对回答的准确性没有太大影响，只有在其中一个问题的回答中，“专业”做法所得到的数据更为准确。

因此，对很多受访者而言，访问员本身就是他们参与调查的关键原因，同时也是决定他们如何进行配合的关键因素。但是，访问员与受访者建立起关系的方式各不相同，而且没有任何显著证据表明哪种交流方式是最好的。

观察人员更加倾向于认为，访问员与受访者之间的交流互动最起码应该具有专业性，它的重点在于调查任务本身，受访者不

应对问答双方的私人关系太过在意。但是,一个访问同时也需要有流畅自由的交流气氛。密歇根和波士顿地区的两项调查都表明,如果访问员能够缓和交流的氛围,对于有些受访者而言是有利的,尤其是那些没有读完高中的人。受访者的喜好和访问员的访问风格显然是各不相同的,对于问答关系的建立而言,没有哪个方式是最好的。不管具体的做法如何,重点在于这一关系中的角色预期,下面我们就来讨论这个话题。

为受访者设立标准

有很多证据表明,访问员如何指导受访者配合访问,对最后的数据质量有很大影响。

为了更好地理解本章内容,我们需要弄清这样一个预设,即在访问之初,受访者是不知道自己应当做什么的;他们并不知晓许多有关调查访问的具体规则。而我们已经说过,标准化访问的特殊之处就在于,访问员要严格按照事先写好的问题提问,除了问题本身以外只能作出很少的解释,而对于回答则要逐字逐句地进行记录。这是一个测量的过程,有着非常结构化的程式,因此受访者也必须了解自己应当如何完成这一任务。而对他们进行指导的就是访问员。

此外,如何评判受访者的表现同样也是个问题。调查访问究竟是一个严肃的测量过程,还是仅仅是填写问卷而已呢?我们所需的是确切的回答还是大致上的泛泛而谈呢?一次礼貌而友善的交谈和一次严肃的社会科学调查,哪个更能带来更高的精确性呢?

有许多因素会对受访者的回答质量产生十分重要的影响。我们在第3章曾经讨论过,访问员对受访者进行指导,教给他们有关标准化访问的一些详细规则,这是十分重要的。而在本章之中,我们讨论的是访问员如何对受访者的表现设立标准。

Cannell 对访问交流的研究中有一项是,在一次访问的过程中究竟会有多少与问答过程无直接关系的行为?答案是令人吃惊的。访问员在访问开始以后所说的话中大约有50%既不是提问,也不是追问(Cannell et al., 1968)。而这些都不是简单的聊天,虽

然有时候访问员会表现得像“登门拜访”一样以缓和气氛；这些都是正式的访问，这些交谈主要还是围绕着访问任务本身的。此外，关于受访者形象及访问员行为的研究也显示，这些交谈与回答质量之间还存在一些有趣的关联，而这在之前的文献中都没有提及。

受访者被问道，他们认为这次健康调查想要了解的是什么。受访者们在一个问题上显示出了很大的差异，这个问题很有意思，即他们觉得访问员是希望他们提供的是“确切的回答”还是“大致的想法”？不仅如此，对这个问题的回答与回答的质量也有着显著的关联。那些认为需要提供“确切回答”的人回答的质量也更高（见表 4.5）。Fowler（1966）发现，在高中以上学历的受访者群体中，这一情况尤其严重。另一个类似的问题问的是“回答需要事无巨细全部都说还是择其要点即可？”。这个问题显示的结果和前一个问题相类似，但是没那么严重。

表 4.5　数据质量与标准化表现的相关性

受访者所理解的访问要求	数据质量指数*				
	低	中	高	合计	N
精确的答案	30%	38	32	100%	223
一般性的想法	38%	39	23	100%	168
事无巨细，都需要	32%	40	29	100%	312
只要回答重要的事情	39%	36	25	100%	80
访问员对精确性的关心					
高	29%	36	35	100%	140
中	39%	39	22	100%	139
低	33%	34	23	100%	133

资料来源：Cannell et al.（1968）。

注：* 这是一个对数据质量进行相对评价的指数，依据受访者的年龄进行了调整。

访问员在访问的过程中所表达出的关切点同样说明了评判标准的重要性。如果访问员在访问过程中经常表露出对精确回答的重视，结果就会得到更好的回答效果（见表 4.5）。同样，这一

现象在高中以上学历的受访者群体中最为突出。

Fowler(1966)把这些结果放到了一起进行分析。他发现,至少在高中以上学历的人群中,访问员在访问中所表露出的关切点会影响到受访者对“确切回答”和“大致回答”的选择。具体来说,重视准确性的访问员会使他们的受访者认为自己需要作出“确切回答”,而重视效率的访问员对“确切回答”的重视程度往往达不到平均水平(如表4.6所示)。

表4.6 高中毕业的受访者对访问标准及访问效率和精确性的认识

	被访问者所报告的访问要求			
访问员对效率的关心	精确答案	概略观点	合　计	N
低	77%	23	100%	62
中	62%	38	100%	81
高	66%	34	100%	47
访问员对精确性的关心				
低	62%	38	100%	104
高	76%	24	100%	86

资料来源:依据 Fowler(1966),第158—159页,表7.3和表7.4,略有调整。

这些数据本身并不能说明什么,但是迄今为止最为重要的发现有许多是在此基础上得出的,它们告诉我们如何提高访问质量以及如何提高以调查为基础的估计的质量。

证明访问员为访问表现设立标准之重要性的较早成果,来自于一项对访问和自填问卷所作的比较研究,问题均与住院有关(Cannell 和 Fowler,1964)。这项研究在当年至少有一人住院的家庭中抽取样本。这些家庭被随机分成了几个不同的小组。对于其中的一组由访问员进行一次健康状况调查,但是没有问及住院事宜。但在访问的最后,受访者会拿到一份有关在此前一年以内家庭成员住院情况的问卷,由他们在访问员离开以后自行填写。而在另一组中,访问员在作完相同的健康调查后,也询问了有关住院的这些问题。

这项研究的权威之处在于他们有一个严格固定的标准来界

定准确性，他们将调查所得的结果与医院记录进行比对，得出的百分比即为准确度。此外，样本在访问员中的分配方式使得他们能够计算出每个访问员所接触的受访者回答的准确度。每个访问员都分配到了不同组别的样本，这样他们所接触的受访者中有的是自填问卷的，也有的是接受访问员提问的。在对结果进行分析后，一个惊人的发现是，同一个访问员的受访者在面对面访问中的问答质量与其自填问卷的回答质量有着很高的相关性。也就是说，如果一位访问员所接触的受访者在面对面的访问中表现很好，那么他们很可能在自填问卷时也表现得很好。

这一发现的重要性在于它显示出了访问员获取高质量回答的方式的意义所在。以往人们关注的焦点在于如何处理问—答环节。然而，很明显，在这项研究中，访问员完全没有参与自填问卷的环节，但是他们在访问中对受访者所设立的标准却在后来自填问卷的过程中对他们产生了影响。这项研究最早证明了设立标准以影响调查数据质量的重要性，也是最有说服力的研究之一。

Cannell 随后更为细致地观察了访问员和受访者之间的互动类型，用以研究访问员如何影响回答的质量。他发现，访问员之间存在的显著差别之一是他们访问的节奏。有的访问员读得非常快，有的则有条不紊且吐字清晰；有的访问员在受访者回答完之后还给他们一点时间继续思考，有的则在得到回答后马上接着问下一个问题。Fowler(1966)发现，访问员进行访问的节奏与受访者对访问的认识有很大关系。如果访问员节奏很快，受访者会认为准确而完整的回答没有迅速的回答来得重要。就我们所知，还没有文件记载表明访问节奏与数据质量之间有什么样的直接关联。但是我们认为，让访问员放慢语速很可能是一个既重要又实用的方式，可用来提高访问中的标准化程度。此外，Cannell 还对四种目前被认为能有效提高回答质量的策略进行了测试。

肯定法(reinforcement)

Cannell 及其同事，特别是 Maquis 等(1972)发现，访问员对受访者行为的肯定有多种类型。在很多情况下，访问员所肯定的会是那些不利于访问的表现。当访问员在访问过程中感到不舒服

时,尤其是那些不情愿的受访者对整个访问都表现出矛盾情绪时,他们的策略就是试图让受访者感觉好些。在不经意之间,访问员会对任何答案都报之以赞许(比如"很好"或是"不错"),不管这些回答是不是经过了认真思考。事实上,即使受访者拒绝回答,访问员也可能会说好话来鼓励他们。

Marquis 设计了一个实验,规定访问员只能对那些有利于访问结果的表现作出肯定。这样,当受访者的回答显得很完整并且是经过认真思考的话,访问员就会对其表现说一些鼓励的话(当然不是关于具体回答的);而当受访者回答得很快或者拒绝回答、显得不假思索的话,访问员就不能作出任何鼓励和肯定。虽然这样做所带来的影响不是很稳定(而且似乎对受教育水平低的人影响更大),但有明显的证据表明,程式化进行的访问与自行进行肯定的访问相比,所得到的有用信息要更多。Cannell, Oksenberg 和 Converse(1977b)也提到了这一点。

显然,程式化肯定方式的困难在于如何建立规范以及如何与恰当的表现一一对应起来。事实上,访问员无法判断受访者的回答是否准确完整,因此选择肯定与否需要有一定的猜测。信息是非直接的,必须根据受访者的状况进行推测。此外,访问员也可能在无意中对错误的回答给出了肯定。Cannell 和他的同事们设计了另外三种方法以对付同样的问题,但得到的结果却大致相似,也许只是在常规的程序基础上更容易应用而已。

模拟法(modeling)

Hensen(1973)认为,要让受访者知道他们应当怎样做,最好的方法就是给他们一个示范。他为一位受访者录制了一卷录音带,其中的受访者对回答的准确性和完整性非常重视。访问员在访问开始时播放这卷录音带,以便"让受访者对访问有一个直观印象",然后才开始进行访问。他们对效果进行了测试,结果发现提供模拟方法确实能够提高回答的质量。

标准化说明法(standardized instructions)

让访问员在访问开始时向受访者宣读一整套说明,这也许是访问员为所有的受访者设立同样的高标准的最直接的方式。Can-

nell 所测试的一套标准话语如下：

"您的回答必须准确并且完整，这一点非常重要。请您不必着急。如果您有任何疑问，请随时向我提问。"

随后访问员开始进行访问。同样，Cannell et al.（1977b）指出这些说明有助于提高回答的质量，并且对受教育程度较高的人群特别有效。

承诺法（commitment）

肯定、参照和说明这样三种对标准化问题的解决方式，还是要靠访问员来完成的，还是由访问员出面说明准确和完整回答的重要性。但是，这三种方式都没有提到受访者的积极作用。它们所关注的只是确保访问员能够成功地传达出正确且一致的信息。Cannell 提出了另一种措施，以便使受访者能够更加积极地参与到好的访问中来。

他所测试的方法如下：当受访者同意接受访问之后，访问员会向他们提出一些问题，"以便使他们大致了解访问会涉及哪些东西"。随后，访问员要求得到受访者的承诺：受访者被告知准确而完整的回答对于研究而言至关重要，他们必须承诺尽量给出最为准确的回答。在最初的测试中，受访者还被要求签署一份承诺书，以保证能够提供最为准确完整的回答。在之后的测试中，受访者被告知，如果他们不能作出承诺，访问将终止。

Cannell 对于拒绝承诺的受访者人数作了充分的心理准备，他原以为被拒率会非常高，但结果并非如此。在最初同意接受访问的人当中，只有 2% 或者 3% 的人中止了访问。而几项测试表明，这样的承诺有助于改善受访者的表现。这也许是最好的方式，即清楚地表明预期并且让受访者积极地认可并遵循这样的预期。

表 4.7 显示的是三种方式——说明、肯定（反馈）和承诺在一次以电话方式进行的健康调查中的综合效果。可以看出，在实验程序中人们所回报的健康状况明显更多（Cannell 等，1987）。在估计中，少报（underreporting）被认为是各种估计中是最主要的误差；有案例认为，估计量越高，其质量可能越高。

表 4.7 不同测试形式的受访者对有关选项的回答百分比

	访问测试的类型	
	控制形式	实验形式
选 项	过去 2 周中,所占百分比为 1 及其以上者	
卧床(Bed days)	7.3%	10.0% *
工作缺勤(Work loss days)	6.3	8.8*
早退(Cut-down days)	8.4	11.5*
看牙医(Dentist visits)	6.8	7.4
看医生(Doctor visits)	17.4	17.5
急诊(Acute conditions)	14.9	17.5*
	过去 12 个月中,所占百分比为 1 及其以上者	
看医生	72.6	74.5
住院日	13.4	12.5*
	所占百分比为 1 及其以上者	
慢性病(Chronic conditions)	29.2	35.8*
限制活动(Limitation of activity)	20.4	27.6*
	个案数(Number)	
近似样本规模 (Approximate size of sample)	4 217	3 993

资料来源:依据 Cannell et al. (1987),略有调整。

注:* 实验综合采取了说明法以及反馈与承诺程序。标* 者,说明在显著性水平 $p < 0.05$ 时,控制与实验之间存在差异。

Cannell 等(1977)所提供的数据更加直观,说明实验性的方法能够改善受访者的表现。例如,在实验条件下,受访者更有可能提供看医生的确切数据,而不是精确到月份或某个时间段,同时他们也会查询医疗记录来确保回答的准确性。

结 论

我们在本章讨论了访问员用以开始访问之方式的三个方面：受访者参与访问的原因，建立问答关系的方式以及评估访问表现的标准。其中，访问员为受访者表现设立标准的方式无疑是造成访问员之间的差异、进而对数据产生影响的最重要的因素，同时也是提高调查的标准化程度、降低调查估计量偏差的最重要的途径。

受访者对自己所参与的访问的了解程度有很大差异。虽然不同的访问员对各自的受访者所提供的背景信息量各不相同，但这些差异对受访者关于访问的了解程度有多大影响，尚无十分确切的记录。我们有理由认为，大多数受访者只需要得到足够的信息来说服自己接受访问就可以了。

一项研究显示，更大的信息量有可能会对那些教育程度较高的受访者的回答质量产生积极的影响（Cannell 等，1986）。而向受访者提供更多有关研究项目的信息也有可能提高数据质量，这一课题值得进一步的研究。在这一点上，我们的实用性建议是，访问员对常见问题要有一套准确一致的回答，同时要留给受访者以自由提问的机会，以便他们在参与访问的时候了解足够的信息。

有关问—答关系质量的建议指南同样是粗略的。虽然访问员在工作时对个性的隐藏有所不同，但也没有一项记录表明哪种访问风格更为优越。虽然对双方互动性的重视可能会对缓和访问气氛有所帮助，但是 Weiss（1968）和 Hensen（1973）的数据表明，注重人际交流的方式反而有可能不利于得到好的访问结果。很可能一个专业的、例行公事型的方式才是一种访问应有的风格。

最后，有很多证据表明，访问员对于准确性的强调存在着显著的差异。几乎可以肯定的是，在访问员之间所有的差异中，这是最能影响数据本身、降低数据质量的一个问题。Cannell 和他的同事们通过改变访问员介绍访问目标以及标准的方式成功地对回答过程施加了影响。结果表明，这是提高标准化程度和数据质量最重要、也最有效的方法之一。

有趣的是，尽管文献资料对这些技巧的记录已经有十多年之

久,但在调查研究的领域内,它们的运用却并不广泛。不得不承认,将应当进行肯定的表现汇总出一个表格实属不易。但是,建立一套标准化的指示,从而在访问开始时就设立好标准,这点并不难做到。此外,让受访者作出一个口头承诺也并非难事,并且已经被证明不论在电话访问还是面对面访问中都是可行并有效的。

标准化的指示、口头承诺以及放慢访问节奏,都是帮助访问员建立标准的直观方式。我们相信,如果将这几项纳入标准化操作程序,那么大多数调查机构都能从中受益,获得更高质量的数据。

第5章

问题设计在标准化访问中的作用

The Role of Question Design in Standardized Interviewing

要确保访问的标准化程度，最重要的方法就是将问题设计得易于提问和回答。在第3章，我们提示给访问员一些技巧以帮助他们用标准化的方式来处理问题。而在第6～8章中我们将要讨论研究人员可以通过哪些方法来确保访问员可以按照要求来操作。在本章，我们主要探讨如何提高问卷本身的质量，因为访问规划得越好，访问的质量也就越高。

在任何调查中，访问员都会对所得到的答案产生不同程度的影响。图5.1中的折线图显示的是130项访问员效应的组内相关系数的分布状况。另外，也有人通过研究发现了类似的结果(Groves 和 Magilavy，1980)。由于访问质量、调查工具及其质量

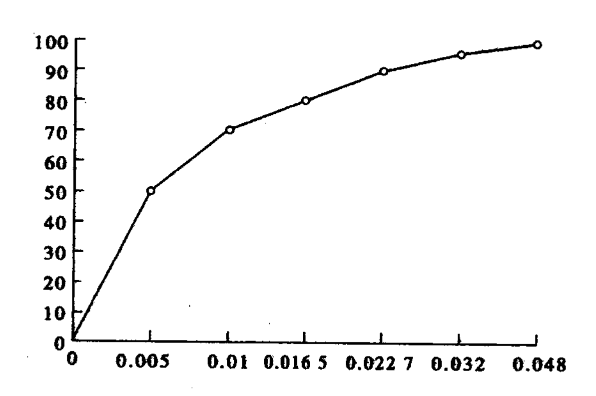


图5.1　130个调查项的Rho系数的累计分布

的不同,不同的研究所得出的结论各不相同,但基本上与我们的结论大致相同。不过,我们现在关心的倒不是这些结果为何会参差不齐,而是为何有些问题相对于其他问题而言更容易受到访问员的影响。

折线越高,表示访问员的影响越大,本章所讨论的重点就是折线最高端部分的一些特征。我们也翻阅了已有文献中所记录的一些发现。我们的结论并不出人意料,因为我们已经说过:那些需要受访者自己作出判断和决定才能获取充分回答的问题,最有可能受到访问员的影响。

问题的内容

表面上看,我们似乎有理由认为,访问员对某一问题的回答的影响程度与这个问题的内容有关,其中有两个方面尤为显著:问题的敏感性以及问题所涉及的是观点还是事件。

资料表明,如果提问涉及敏感话题或有难言之隐,受访者在回答时就会有所保留。比如 Locander 等人(1976)就提出,人们一般不太会提供有关自己破产和酒后驾车被拘的情况,而其原因显然不可能是人们遗忘了这类事情。Cannell 等(1965d, 1965a, 1977a)也指出,与一般的住院情况相比,人们一般不太会提供那些由于难言之疾或是疑难杂症而住院的经历。但是,我们关注的重点并不是这些;我们所关心的不是关于这类事件的回答有多么准确,而是不同的访问员在测量过程中的一致性如何。我们的确有理由认为,有些访问员所建立起来的问—答关系与其他人相比更有助于对一些敏感内容的坦陈。但是,能够支撑这一假设的经验证据却寥寥无几。

我们所看到的论据大多来自 Sudman, Bradbrun, Blair 和 Stocking(1977)。他们发现,访问员对于一个问题会不会使受访者感到麻烦或敏感的预期,关系到他们是否能够得到充分的回答。如果访问员认为一个问题对于受访者而言很敏感,那么他有可能根本不会得到任何回答。但是,对于已经得到的回答,Sudman 并没有发现访问员影响与他们对问题敏感性的预期有什么显著的关系。

在我们的研究中，我们将一系列问题按照回答的敏感程度进行编码，并观察了其与内部相关性的联系。我们没有发现敏感性与内部相关性有明显的关系；事实上，敏感话题的内部相关性甚至要更低一点。因此，虽然这一理论看起来很有说服力，但没有任何研究证据能够证明敏感性内容在使用标准化方式进行测量时的难度会大于对非敏感性内容的测量。

Kish（1962）认为，询问观点或态度的问题比询问事实情况的问题更容易受到访问员的影响。他的论据是，事实型问题（可以被客观识别的过往事件或当前状况）所依据的是客观事实；而观点型问题则相反，除了受访者本人以外没有人能作出观察和界定。有人认为，访问员引起的误差不外乎两种，要么是受访者在其影响下改变回答，要么就是受访者为了迎合访问员的预期而主动改变回答。因此，这一假设认为观点型问题比事实型问题更容易产生这种变动。

虽然这个假设看起来很有道理，但同样缺乏数据支持。Kish的数据并没有显示这两种问题在访问员影响方面有任何不同。我们的研究同样没有发现访问员对这两种问题的影响有何差异。总的来说，我们发现事实型问题受访问员影响的程度与观点型问题相比甚至还要略高一点。

但是，并不是所有的观点型问题都是这样。那些非常难以回答的观点型问题，通常是那些受访者对于这些问题不太可能有完备的思考的问题，就明显受到访问员的影响。但是，我们十分确定，使回答受到访问员影响的并不是问题的内容，而是这些问题的其他特征。下面我们就要讨论这些特征。

总的来说，还没有研究文献能够证明，问题的内容与问题是否易于标准化有什么固定的联系。

问题的形式

问题在形式上最明显的差别之一就在于它究竟是给出几个答案供选择（称之为封闭型或固定选项问题），还是让受访者用自己的话来回答（称为开放性问题）。后一种类型的问题更有可能受到访问员的影响，因为什么样的回答符合要求并没有明晰的界

定,而且访问员在记录答案时也更加困难。

在研究文献的记录中,有几种访问员的人为误差主要是针对开放性问题的。例如,Hyman 等(1954)提出,访问员会根据自己对受访者"真实处境"的判断而选择性地对不完善的回答进行追问。在答案的记录上,也同样发现了误差,访问员会遗漏掉一些他们认为与受访者的情况不符的要点。而对于封闭型问题,则不存在以上这两个问题。

在研究内部相关性时,Groves 和 Magilavy(1980)发现,有效回答的数目以及得不到有效回答的可能性,都与访问员有很大关系。同样,这只是开放性问题的情况。

但是在我们的研究中,我们大量对比了开放性问题和封闭型问题的内部相关性,结果没有发现显著的差异。虽然有证据表明,开放性问题更有可能带来问题,但是并没有统计数据证明这是一个必然趋势。

问—答过程

我们无法仅仅通过表面特征来判断一个问题作为标准化测量工具的优劣,因此我们还要深入研究问答互动的具体细节。我们从 57 位访问员所做的访问中选取了 100 份录音记录,并且将问—答过程中所发生的情形作了详细的编码。我们挑选出 8 种有可能对测量过程产生影响的行为,并且计算了它们发生的频率。这些行为是:

1. 访问员是否严格按照问卷卷面的措辞来表述问题;
2. 访问员进行追问的方式是否正确;
3. 访问员的追问是否带有导向性;
4. 访问员是否对所有需要追问的问题都进行了追问;
5. (仅针对开放性问题)访问员是否逐字逐句记录了回答;
6. 访问员是否对受访者的回答作出了不恰当的反馈;
7. 访问员是否作出了不恰当的个人交流行为;
8. 问答的过程中是否有笑声。

我们计算了每次访问交流中上述行为发生的次数,并进行了

仔细的编码；之前我们曾经根据组内相关系数（intraclass correlation coeficient）计算出了哪些问题会或者不会带来显著的访问员影响；我们将这些访问员行为编码后与之进行对比，以弄清二者之间有何关系。表 5.1 列出了比对的结果。

表 5.1 访问员行为与 *Rho* 系数的显著相关性

访问员行为	与 Rho 的相关性
笑	—
朗读问题失误	—
正确进行追问	.23
诱导性追问	.20
追问失败	.49
对答案的反馈不合适	—
人际交流行为不合适	—
没有完整、准确逐字逐句记录开放性问题的答案	.39

注：标破折号的单元格是在 $p < .05$ 时，相关性不显著。计算相关性的依据是对 65 项问题进行 100 次不同访问时进行的行为记录。自由度 =64。

不恰当的回馈与个人交流行为从各个问题的平均值来看，发生的频率并不很高，这也可以解释为什么它没有对单个问题的回答产生影响。这不是说人际交流行为不会对数据质量产生影响，而是说明在训练有素的访问员中，不能用这些行为来解释为什么某些问题有更多的可能或者更少的可能受到访问员的影响。

我们原本以为，被访问员误读或改动过的问题是最容易受到影响的；但事实并不是这样。对于所有的问题而言，访问员是否严格按照字面所写的去读与访问员影响的大小并没有什么联系。

但我们发现，交流过程的其他几个方面确与访问员影响存在着联系。最主要的影响因素是，一个问题在多大程度上需要进行追问。上述 8 个行为中与追问有关的三个方面——正确追问的数目、引导性追问的次数以及应当追问却没有追问的情形，都会对内部相关性产生影响。简单地说，如果有一个问题需要进行追问，那么访问员就多了一个出错的可能性；只要有一个问题需要追问，访问员就有可能带有倾向性，就有可能对需要进行追问的

问题没有作出应有的反应；由此而对数据产生的影响都属于访问员影响。

另一个与访问员有关的组内相关，是访问员在记录答案时所产生的误差。虽然这类记录误差没有追问误差来得普遍，但我们仍发现，访问员有选择地记录回答，是降低测量过程标准化的重要原因。这一点与 Hyman 等(1954)的结论是一致的。

我们查阅了已有的研究文献，并且细致分析了我们自己的研究中有关问答互动的部分，试图找出需要进行追问的这部分问题的特征，以及是否有某些情形特别容易导致访问员出现错误影响。我们认为这是一个重要的领域，有待进一步的研究，而我们目前的结论还很不明晰。但我们发现，问题的某些特征有可能会使访问员对数据产生影响。

与追问有关的问题

我们对于问题特征的研究，没有发现有哪种类型的问题会比其他问题更容易受到访问员的影响，但是有几类问题更有可能需要进行追问，因而也就更有可能受到访问员的影响(见表5.2)。

表5.2 访问员行为与问题特征的显著相关性

	问题特征			
访问员行为	难度	敏感性	观点	开放题
笑	.34	—	.31	.35
朗读问题失误	.52	—	.45	—
正确进行追问	.34	—	.24	.63
诱导性追问	.59	—	.57	.48
追问失败	.28	—.22	.38	.56
对答案的反馈不合适	.29	—	.27	.42
人际交流行为不合适	.28	—	.24	.23
没有完整、准确逐字逐句记录开放性问题的答案	.28	—.26	.38	.59

注：标破折号的单元格是在 $p<.05$ 时，相关性不显著。计算相关性的依据是对65项问题进行100次不同访问时进行的行为记录。自由度=64。

具体而言,开放性问题比封闭性问题更有可能需要追问;观点型问题比事实型问题更有可能需要追问;回答难度高于平均水平的问题(可能是因为需要回忆一些比较难于记忆的东西,或是因为受访者此前从未对相关问题有过任何思考或想法),也更有可能需要追问。此外,高难度问题和观点型问题也更有可能导致引导性追问。

因此,我们的结论并不是说这些类型的问题(高难度、开放性或观点型问题)本身容易受到访问员的影响;而是这些类型的问题需要追问的可能性明显更大,因此更不容易使用标准化的方式进行测量。

标准化测量面临的若干威胁

通过录音并对问答互动进行编码的方式可以准确地鉴别出哪些问题需要进行追问。但是,通过简单的预调查的方式我们还发现了至少五种问题特性会使问题本身更容易受到显著的访问员影响。它们是:

1. 所提及要点的数目;
2. 带有潜在预设的问题;
3. 问题中的概念不清;
4. 对访问员的要求不明晰;
5. 需要进行现场编码的回答。

要点数目

通常在回答询问观点的开放性问题时,受访者在回答中所涉及要点的数目会对数据产生影响。

例:住在这个小区让你感到最喜欢的是什么事情(things)?

这个问题允许受访者给出多个回答。事实上,复数名词事情("things")会让访问员认为他们至少应该追问一次来获取第二个回答。但是,这个问题本身并没有对访问员和受访

者明确到底回答几次、提到几个要点才可以。正如此前所述,Groves 和 Magilavy(1980)指出,受访者回答中的要点数目最容易受到访问员影响。如果研究人员没有作出明确要求,访问员之间就会出现不一致的情况。

带有潜在预设的问题

以下是几个带有潜在预设的问题:

1. 在过去的一年中您读过的书中最好的是哪本?
2. 下次选举您会投谁的票——Harry Truman 还是 Thomas Dewey?
3. 昨天在您身上发生的最好的、最让您高兴的一件事是什么?

调查研究人员提出的许多问题,也包括以上三个问题,实际上其中都同时包括了至少两个问题。其中一个潜在预设的问题,实际上就是提出了对受访者而言适用与否的问题。

分析:

第一个问题假定了受访者在过去的一年中至少读了一本书,(而两本以上则更有利于回答这种比较型问题)。受访者可能会觉得这个问题对他而言没有任何意义,因为他一本书也没读;受访者也有可能拒绝回答这一问题,因为在读过的书中他一本都不喜欢。虽然在理论上讲人们可以从一堆不喜欢的书中挑出一本相对最喜欢的,但是我们的确有理由质疑在这种情形下这些问题还是否有意义。

关于投票的第二个问题所包含的潜在问题是,受访者是否打算去投票;更重要的是,受访者是否已经决定或是有意向支持哪位候选人。所以这个问题事实上包含了三个问题:受访者是否打算投票,他是否已经在两个候选人中作出了选择,如果是,那么他支持的是哪一个。

最后一个问题,如果咬文嚼字地来分析其实也没有错:不论昨天发生的事情有多可怕,总有一个是最好的。但是,如果受访者认为昨天发生的事情没有一件是让他高兴的,那么他有可能回

答说“没有”。同样,这个问题事实上被当成是两个问题:“昨天发生的事情中有没有让您感到高兴的?”如果有,那么是“哪件事?”

这些问题都是“不知道”型问题的变种。对于几乎所有的问题而言,受访者都有可能会说他对此不了解,或是他对这个问题没有足够的思考,因此不能给出恰当的回答。

从标准化的角度来看,棘手之处在于问题本身并没有注明任何处理这种隐含问题的规则或程序(procedure)。这些隐含的问题并没有被明确提出来,但有的受访者却能够揣测出问题的隐含内容,并对此作出回答。如果受访者主动对隐含问题作出回答,访问员就需要决定如何处理这一情况。而这种情况的出现其实就已经说明,原先设计的这个问题在此并不适用。如果事先没有明确指示,访问员在处理这些情形时就会采用不一致的方式。研究表明,访问员之间接受诸如“不知道”或“对我不适用”之类的回答的比例有显著的差异。此外,Schuman 和 Presser(1981)也指出,对“不知道”型回答的处理对回答的分布(distribution)有着实质性(substantive)影响。

概念定义不清

通过研究我们还发现,含有模糊概念的问题显然更容易受到访问员的影响。我们将问题按照是否含有“模糊”概念进行编码。虽然此类问题的样本容量没有理想中的那样大,但我们仍没有发现任何证据表明这类问题与访问员所引起的内部相关性的大小有任何关系。

尽管如此,在我们对问答双方的互动进行深入细致的研究后,我们发现,对关键概念进行解释时所发生的非标准化对话、交谈会导致显著的访问员影响。由此,我们的结论就是,只有某些类型的模糊概念会导致非标准化问题。

一个重要的发现是,模糊概念并不一定会引发追问。有些概念和问题的确很模糊,表意不清,但是对受访者而言这些都不是问题。例如,受访者被问到在过去的一周内有几天吃了早餐。我们可以肯定的是,每个受访者对早餐的定义是各不相同的。但是,受访者都认为自己能够理解“早餐”的意思,无论这一理解与

其他受访者或是研究者是否相同。因此,虽然从测量的角度讲这并不是一个明确的概念,但它并不会导致访问员与受访者之间的非标准化交流。

我们认为,一个带有模糊概念的问题如果能够通过重复问题的方法得到恰当的处理,也很可能实现标准化测量。如果受访者、访问员以及研究者对问题有相同的理解当然更好;表意不清的问题虽会导致误差,但是如果它没有在访问员和受访者之间引起交流,没有给访问员表现不一致的可能性的话,这个误差就与访问员没有关联。如果访问员知道当受访者对表意不清的问题提出疑问时如何处理,那么就不会产生不一致的情况。

当访问员不是仅仅重读一遍问题或是在被要求解释概念的时候说"按照您自己的理解"就能解决问题的时候,那才是真正的难点所在。

以下两个问题就属于上述情形,我们的研究显示,它们受到了访问员的强烈影响。表意不清的概念用斜体字标注。

1. 您上一次*看医生*是多久之前——上个月,1 至 6 个月之前,6 至 12 个月之前,还是一年多以前?
2. 在过去 12 个月中您是否患过*湿疹*或*牛皮癣*?

问题 1 的模糊之处在于"看医生"究竟指什么;而在问题 2 中,湿疹和牛皮癣这两个词对于受访者而言并不熟悉。

这两个问题本身都没有对其中不明晰的地方作出解释,因此当受访者提出疑问的时候,访问员显然不能置之不理。于是,在访问员和受访者之间就会发生非标准化的交流,从而导致严重的访问员影响。

对受访者的要求不明确

在表意不清的情形中有一种特别重要的类型也会导致访问员影响,那就是问题本身没有说明怎样的回答才算是充分回答。从某种意义上说(in one sense),这一问题的一个特殊例子就是上文曾经提到的,对开放性问题的回答中所提及的要点数目。但是,通过对问一答环节的研究,我们发现此类问题还有其他一些更为普遍的情形。

"您认为您主要是通过什么渠道来获取有关健康以及如何保持健康的信息的?"

虽然从表面来看这好像是个合乎情理的问题,但它其实对受访者和访问员提出了两个问题。第一,所谓"渠道"是指什么。人们通过不同的方式获得信息,比如与人交谈、阅读或看电视,这些是否都算是渠道,还是说它们都只算是通过另一渠道获得的信息?

与此相联系的另一个问题是,回答需要有多详细。比如,"通过阅读"算不算是个充分的回答。如果不算,关于阅读的种类是否需要更详细的说明(如杂志、书本、报纸),甚至是不是需要细化到文章的出处呢(如《新英格兰医学杂志》的论文,*Ann Landers*,《星座》(*my horoscope*)等)。

问卷的字面表述本身绝对不能解决这些问题。结果是访问员需要作出自己的判断,因此不一致的情形就会很频繁地出现。首先,访问员要自行判断一个回答是否足够详细具体;如果认为回答不够具体,他们事实上也没有什么实质性的工具来帮助受访者回答,他们只能按照自己的想法通过进一步提问来解决这个问题。

实地编码的问题

我们确实不提倡这种实地编码型的问题。这种问题以开放性问题的形式提出,由访问员将所得回答归入给出的几个大类。这些问题要求访问员进行实地编码。但访问员并不是按编码员的要求进行筛选和培训的,除了编码之外,他们还有其他许多任务要做;最糟糕的是,他们的编码结果无法通过检查编码来检验出误差及他们在编码时存在的欠缺。

我们的研究仅包含了少量实地编码型问题。从平均值上看,这种问题与其他问题相比引起访问员效应的可能性并没有大很多。不过,情况最糟糕的两个问题恰恰属于这一类型。其中一个原封不动地选自美国健康普查。"您上一次去看医生是出于什么原因?"访问员需要将回答归入下列几个类型之一:

1. 诊断或治疗
2. 体检

3. 预查或复查
4. 眼部检查
5. 预防接种
6. 其他（请说明）

另一个实地编码型问题则是我们自己想出来的。"在您的生活中您最想发生变化的是哪一方面？"访问员需要将回答归入下列几个类型之一：

1. 工作
2. 孩子
3. 财务（finance），钱
4. 健康
5. 闲暇
6. 其他（请说明）

这些选项（item）的内部相关性很高。不仅如此，我们对问一答互动过程进行回顾之后发现，它们还会导致访问过程中最糟糕（即标准化程度最低，最具引导性）的一些行为。

原因在于我们赋予访问员一个根本无法完成的任务。一方面，这两个问题都没有明示什么样的回答才算充分。询问"为什么"的问题通常都很麻烦。可以设想对"为什么去看医生"的几个可能的回答："我先生让我去的"或是"因为我信任医生"。尽管许多受访者显然都知道在这种情况下要问的是与医疗健康有关的原因，但是问题的字面表述本身并没有明确这一点。

而真正的问题在于，受访者对访问员的要求一无所知，而访问员也不知道如何以一种标准化的方式对受访者进行指导。受访者会给出一个相对比较符合某一类型的回答，但这种归类并不明确。访问员当然知道这些回答应当归到哪一类，但是他们在追问时遵循非引导性原则的严谨程度不一样。此外，从录音回放中可以听出当访问员没能得到一个明确的回答的时候他们有多郁闷。最后，一些访问员可能会问："那么您觉得这算是体检，对吗？"

没有明显的证据表明实地编码型问题在任何情况下都会带

来很糟糕的结果。只是在某些情形之下会有糟糕的结果。具体而言，如果问题本身对回答的特征和详细程度都作出了明确的要求，那么即使它属于范围编码型问题也不会有影响。但是，如果受访者从问题本身无法获知其对回答提出了哪些要求的话，这样的问题就很失败。

也许关键不在于问题是否属于实地编码的类型，而是开放性问题如果没有对回答作出明确要求的话就会导致访问员影响。导致访问员影响的也可能不是范围编码本身，而是问题的其他方面。编码范畴本身的质量，如明确度、非交叉性、对不同回答的包容性及如何表述，都会影响到访问员的表现。但不管怎样，在设计标准化访问表格时，我们对于实地编码型问题的使用仍然持保守态度。

设计调查工具以使访问员效应最小化

以最小化访问员效应为宗旨来设计问题，归结起来就是在问题本身加入一些信息来指导受访者进行回答。在设计问题时将所有访问员需要说的话都包括进来，这样当需要对问题进行解释或追问的时候他们就不会作出五花八门的决定。

这听起来也许算不上有多深刻。读过 Stanley Payne(1951)的书的人可能会认为这些原则在书中早就有所提及。但是，目前的调查中仍然有相当多的问题在编写时并没有遵循这些标准，因此访问员处理问题的方式也各不相同。此外，我们认为，在设计问题时应将追问的必要性降到最低，同时还要对回答问题时所需要的额外信息作出预计。但是，从我们与同行们在问题设计上的工作经验来看，对这方面的重视程度还是远远不够的。

我们对于问题缺陷的讨论得出了几个有助于完善问题设计的法则。首先，要在问题中加入对关键概念的解释，尤其是当需要提供某类事件发生的次数的时候。不要把缺乏明确定义的问题丢给访问员和受访者。

第二，将带有潜在预设的问题表述得更加明晰，这样无疑会带来更完善、更加标准化的测量结果。虽然 Schuman 和 Presser (1981)没有明确要求优化具有潜在预设性的问题，但他们的数据

在我们看来却是具有强制性的。我们不应当让访问员或受访者以各自不同的方式来决定某一问题对自己是否适用,相反,我们应当设计一个统一的程序来让受访者决定某一问题对自己是否适用。这样的设计能够降低由于不同的受访者中存在着差异而导致的误差。

第三,研究人员如果在编写问题时明确提出对受访者的要求,则可以减少由访问员引起的误差。一个显著的方法就是使用封闭性问题而不是开放性问题。在对何为恰当的回答进行界定时,最清晰的表述莫过于"请从中选择一个"。

在使用开放性问题时,研究人员应当想一想怎样才可以缩小回答可能的涉及面。要避免使用怎样(how)、什么时候(when)、什么地方(where)、为什么(why)、多少(how mang)这些副词,这样可以更加明确地表述出特定的问题需要什么样的回答,否则,访问员往往会需要通过改动问题来获得符合要求的回答(见表5.3)。

表5.3 副词的等同语(Adverbial Equivalents)

副词式表述	替代解释
1. 你是如何来上班的?	a. 来上班时,你搭乘什么交通工具? b. 来上班时,你走的哪个路线?
2. 你是什么时候搬到这个地方的?	a. 你是哪年……? b. 多少年之前你……? c. 你多大年纪……?
3. 为什么您投票支持候选人X?	a. 候选人X的什么特点吸引了您的选票? b. 候选人Y的什么方面使您不愿意支持他? c. 您投票支持候选人X,是基于哪个方面的利益或考虑?
4. 您十多岁的时候住在哪里?	a. 您住在哪个城市或者城镇? b. 您和谁住在一起? c. 您住在什么样的房子里?
5. 您挣多少钱?	a. 您获得了多少美元报酬? b. 与其他人相比,您的收入如何? c. 与您的需要相比,您的收入如何?

第四,对于开放性问题的回答所涉及的要点数目,不同的访

问员和受访者之间会存在着差异。对此可以通过明确要点数目的方式进行控制。我们倾向于询问“主要原因”而不是“所有原因”,这样每位受访者就只给出一个回答。如果实在需要多项回答,我们可以额外询问次重要的原因。我们的目标是消除访问员和受访者对于测量任务的不确定性。这样,决定受访者如何回答问题的就是研究人员,而非各个访问员或受访者。

不过,比这些改进问题设计的建议更为重要的是,我们需要在预调查或研究的试验性阶段对问题作出仔细的评估。我们的研究表明,我们不能想当然地鉴别出哪些问题可能会导致严重的访问员效应。那些需要通过不断追问才能获得合格回答的问题,尤其是那些无法仅仅通过重复问题的方式得到解决的问题,最容易受到访问员的影响。在这些问题中,有封闭型问题也有开放性问题,有观点型问题也有事实型问题。但是,这些问题中有很大一部分能够在研究的试调查阶段被鉴别出来。我们认为,在问卷最终定稿之前鉴别出这些极有可能需要追问的问题并加以改进,是减少访问员误差的最有效的方法之一。

有待改进问题的识别策略

一个典型的试调查要求有调查经验的访问员进行 10 到 20 次访问,然后与研究人员一起开一个 1～2 小时的会议来讨论他们遇到的所有问题。这种试调查是很有用的,但我们认为,研究人员需要做的还远不止此。我们和其他同行正在设计和评估更好的方法来检验问题。DeMaio(1983)以及 Converse 和 Presser(1986)对目前评估问题的方法分别作了一个更为详尽的回顾。我们认为以下四个步骤都十分有助于问题设计的改进:

1. **焦点讨论小组**是设计调查工具之初最古老也是最有效的一个方法。在被访问的人群中选取 6～8 个典型的受访者组成一个小组,通常有 2～4 个小组就够了。这些焦点小组针对访问所涉及的范围就各自的经验和想法进行讨论。通过使用便宜的录影带,整个研究小组可以在不出席讨论的情况下对整个讨论过程中所说的话进行回放。

从降低访问员效应的角度而言,这种讨论的具体作用包括鉴别出问题中的表意不清、需要明确定义的概念,明确某些问题所要求的回答范围,以及辨别出问题中可能夹带的关于受访者应当如何思考、如何作为的错误暗示。这种焦点小组讨论不仅可以帮助研究人员设计出一个标准化程度更高的调查工具,而且很有可能帮助他们提出更有助于收集有用信息的问题。我们相信,在设计调查工具之前组织焦点小组讨论对所有的调查研究项目而言都将是有所裨益的。

2. **认知研究技术**(Cognitive research techniques)最近才被引入调查研究领域(如 Jabine 等,1984)。长久以来,认知心理学家一直在研究人们如何处理信息,如何回忆信息以及如何进行思考。但这些技术直到最近才被用来在调查中设计更好的问题。他们的方法主要是对几个比较典型的受访者进行相对高强度、长时段的访谈。目前,看来使用最为广泛的技巧是所谓的"出声思考"式访谈。在这样的访谈中,受访者需要在接受试调查的过程中将自己的思考过程说出来;通过听取人们对问题的理解和思考,研究人员得以获知问题表述的模糊之处以及受访者回答时的难点所在。

   这种方法的一个变体是将访问过程重复进行两遍。首先,受访者像接受普通的试调查那样回答问题;随后,访问员带领受访者将整个访问过程再回顾一遍,让受访者对作出的回答以及遇到的问题进行讨论和解释。

   这些技巧仍在不断地发展和评估之中。这需要花费大量的人力,但却可以有效地鉴别出可能导致误差的问题。

3. **更好地运用试调查**也十分有用。尽管有经验的访问员肯定有能力鉴别出问题中存在着的缺陷,但通过观察我们发现,不同的访问员对问题的界定是有很大差异的。事实上,有经验的访问员在处理由于问卷设计缺陷而导致的问题时显得更为妥当;我们还发现,对于这些缺陷所带来的一些困难,他们也会表现得并不敏感。

在试调查过程中，我们曾经尝试发现更多的问题，以便对访问员进行更多的培训。我们的目标是提高他们识别那些读起来拗口的问题以及通常需要追问的问题的敏锐度。我们还认为，让访问员对每个问题进行标准化的书面评分，能够使他们的输入更为系统和有用。我们仍在探讨让访问员对问题进行评估的替代方式，但我们深信，更加关注访问员的培训，关注任务简报的过程(debriefing process)，就能够提高访问员鉴别问题缺陷的能力。

4. **录音记录和编码试调查**过程，在我们看来是评估调查问题时最重要的一个创新。为了有效使用这一方法，至少应当进行 25 次试调查访问，以提高测量的稳定性。访问员在征得受访者同意后对访问过程进行录音。无论是面对面访谈还是电话访问都可以使用这一方法。在电话访问中，当访问员征求受访者同意的时候，一定要将对方的回答录下来。

对录音的编码工作需要大约两倍于访谈本身的时长。我们所进行的编码工作是让一个编码员对四五种不同的行为在每个问题回答过程中出现的次数进行计数：访问员是否准确按照书面表述来提问，受访者是否要求作出解释，受访者的回答是否充分，是否需要追问。对于最后一个问题我们使用了两种方法：

A. 编码员对访问员进行追问或重复问题的情形进行计数；
B. 编码员对受访者回答不充分或不完整的情形进行计数。

我们计算了这些情形发生的比例，即问题被误读的比例，受访者要求解释的比例，问题需要额外追问或导致不充分回答的比例。

我们仍在努力寻求最好的编码方式，以实现结果的稳定性，并且以尽可能低的成本运用它来检测问题的质量。但是，之前本章已有数据显示，我们有证据表明，需要额外进行追问的问题显然更易受到访问员的影响。同时，我们还很确定这项技巧能够有效识别出那些读起来很拗口的问题。尽管我们有望通过更深入的工作来记录下这些措施的价值，但是到目前为止我们还是相

信,对问答互动的过程进行编码,并对那些妨碍标准化访问的因素给予特别的关注,能够有效、可靠地帮助我们识别问题中存在的缺陷,从而在调查全面铺开之前对其作出改进。

结 论

在接下来的三章中,我们将要讨论怎样使访问员在访问过程中按照我们的要求去做。但是,我们将会看到,研究人员控制访问员的能力是有限的。

正如我们在本章之初所提到的,在已出版的有关访问员对数据影响的研究中,有 1/4 到 1/3 的问题由于受到明显的访问员影响而产生了误差。此外,几乎可以肯定的是,资料文献中所使用的研究案例在同类研究中无疑处在较高水平,他们对访问员的培训以及对调查方法的选取都更为重视。

虽然挑选那些有能力并且愿意去作标准化、非引导性访问的人作为访问员,其本身也是减少误差的一个方面,但他们所分配到的任务同样会对标准化的目标能否实现有着至关重要的影响。通过研究我们发现,有些问题对访问员和受访者提出的是一些根本无法完成的任务;使用问题所提供的词汇完全不能得到预期的回答。在这种情况下,访问员别无选择,只能自作主张;有时他们只需要自行决定使用何种方式进行追问,追问几次;有时则需要彻底改写问题,对问题的含义作出自己的解释,并指导受访者进行回答;不论他们需要做些什么,一旦访问员开始采取这些行为,他们所使用的行为方式就是千差万别的,结果必然会导致访问员人为误差。正如前人总结的,最重要的是 Bradburn 和 Sudman (1979)指出的,降低访问员效应的最好方法是将问题编得更好;而更好地设计问题的关键在于在调查开始前进行细致、彻底的试调查。

访问员选择及访问员相关误差

Interviewer Selection and Interviewer-Re;ated Error

第6章

我们挑选访问员的时候应该注重哪些特质呢？对于这个问题，我们的关注点在于，访问员的这些特质与其收集到的数据是否有联系？如果有，那么联系有多密切？

我们认为，访问员的某些特质有可能对采集的数据产生影响，这样说的依据有三个。首先，有些访问员比其他人要更善于驾驭问答流程；第二，访问员的某些特质有可能改变问题的语境或含义；第三，访问员的特质会对问答双方关系的质量产生影响。我们将在本章证明这三点。

访问员特质及绩效

有些特质是决定一个访问员优秀与否的必要条件，如果说尚不是充分条件的话。其中最显著的一条就是良好的读写技能。在美国，被雇用为专业访问员的人几乎没有高中以下学历的。尽管有些调查机构特别要求访问员具备大学以上学历，但是我们目前还未发现有任何研究对访问员受教育程度与其收集的数据质量间的关系作过系统考察。

Sudman 和 Bradburn(1974)进行了一个广泛的统合分析(meta-analysis)，考察访问工作人员的特质与他们所收集数据的质量之间的联系。他们发现的唯一规律是，使用年轻人作为访问员的研究中存在的误差更多。Sudman 确信其原因在于访问员，因为雇用大学生做访问员的调查通常都不太正式，而且这些访问员往往没有受过什么训练，此前也没有访问的经验。

有许多研究试图找出工作人员与调查问卷对最后结果的影响。虽然个别研究发现二者之间存在着一定的关系,但我们在已有文献中并没有发现任何记录表明访问员的性格特征(不论使用何种标准测量)与对问答流程的驾驭能力之间存在任何关系。

一个比较常见的问题是,是否应当要求访问员受过与研究领域相关的特殊训练。当需要对某一领域内的专业人士进行访问时,这一问题就显得尤为重要。例如,人们会要求由法律系的学生来访问律师,由护士或医学系的学生来进行健康调查,或者更笼统地讲,健康调查要由护士来做。

在此我们想把"如何有效地寻求配合"和"如何有效地执行(conduct)访问"作为两个问题加以分别讨论。有一些零散的、非正式的(anecdotal and uncontrolled)证据表明,如果访问员看起来像是同行,受访者表示配合的几率就会更高。比如,这些证据表明,在对警务人员进行研究时,如果访问员有过执法经验,则受访者配合访问的可能性会更大。但是,从所收集的数据质量来看,我们没有发现任何依据证明专业背景能够帮助访问员更好地完成一项标准化访问。事实上,通常的情况是,有没有专业背景其实无关紧要,而许多有经验的调查研究者反而认为,没有任何专业背景或是未受过某一领域专业训练的专职访问员,对访问的把握会更加妥当。

我们很难证明这一结论,因为针对具备特殊专业背景的访问员所进行的培训和管理与普通访问员通常是不一样的。但是,似乎存在这样的情形,比如法律系学生在对律师进行访问时,他们不太会对模棱两可的回答进行追问,而是会装作已经理解了受访者的意思,即便这些回答并不明晰。这可能是因为有些行话虽然表意不清但在他们看来都很熟悉,也可能是因为他们即便不能完全理解一个问答的时候也不愿意去追问。如果访问员错误地认为自己已经理解了受访者所说的话,就不会进行追问或是进行引导性的追问,从而产生误差。而一个普通的访问员则不具备这些专业领域的知识和技能,因此可以没有顾虑地进行非引导性追问以获得一个明确的回答。

有时,调查人员会花两天的时间来收集数据,通过一些额外的观察和分类对已有数据进行补充和完善,而这些通常需要求助

于专业知识背景。例如，在调查老年人对援助服务（support services）的需求状况时，有可能需要进行一些临床评估；而护士显然更有能力做好这一工作。我们的经验是，如果要通过观察获得一个标准化的测量，研究人员最好能设计一个任何合格的访问员都能使用的测量方法。而且，事实证明，普通的专业访问员经过训练后也能够准确地测量血压，检查身体状况，以及许多访问以外的事情（如 McKinlay 等，1982）。

也许，人们选择具有专业背景的人作为访问员，除了提高访问质量外还有其他原因。但是，就访问收集的数据质量而言，我们并没有理由要求访问员具备特定领域的专业背景。

访问员特质与访问情境

标准化数据收集的目标是让访问员遵循统一的操作方式。虽然我们可以训练访问员不要改动问题，不要用引导性的方式进行追问，但是在访问中他们显然不可能将自己的个体特征完全隐匿起来。在面对面的访问中，受访者可以一眼看出访问员的年龄、性别和种族背景，还可以进一步推测出他们的教育水平、社会阶层、民族和宗教背景等特征。在电话访问中，受访者的观察相对局限一些，但性别是很好判断的，而且从访问员的姓名、口音、语气以及语言风格可以判断出一些信息，不管其准确与否。

从操作的角度来讲，研究人员不可能找出两个各方面都一模一样的访问员来。此外，人们一般都认为访问员和受访者之间的特征差异才更为重要。因此，一个 45 岁的访问员在对 20 岁、45 岁或 75 岁的人进行访问时所起的作用是不同的。

这样的话，测量的标准化就只能指望访问员的这些特征不会与所得答案相关，不会对获得的回答产生影响。我们完全有理由认为，访问员的年龄、种族或性别对大多数问题而言是无关紧要的；而且事实的确如此。尽管 40 多年来，人们一直在研究访问员个体特征与访谈回答之间的关系，但是很少有实例能找出这二者之间的联系。

但是，在对这些发现进行总结之前，需要指出的是，我们很难将访问员所有的个体特征都纳入研究的范围。美国的社会科学

研究课题所雇用的绝大多数是受过四年以上高中教育(high school education)的白人女性。而如果要研究访问员的种族、性别、教育对回答的影响,则必须要将一群特征各异的访问员集中到一起来分析。新人的表现往往不如经验丰富的访问员。后者都是经历了长期的磨合筛选后优胜劣汰的结果。除此之外,男性、小团体成员或是高中以下学历的访问员相对而言是不多见的,即使有,他们之间的差异也有可能很难明确界定。由于这样那样的原因,有时很难将源自访问员个体特征的影响同经验、训练、兴趣这些因素的影响区分开来。

宗教或族群

最早关注这一问题的一项研究细致分析了具有明显犹太特征的访问员在收集可能带有反犹太主义倾向的回答时所产生的影响。访问员被分为三组:第一组访问员的姓氏及相貌都带有明显的犹太人特征;第二组访问员的姓氏带有犹太特征,但相貌上没有明显特征;第三组访问员在姓氏和相貌上都无明显的犹太特征。这三组访问员随后对三组对照样本进行访问。尽管多项指标并未受到访问员之犹太特征的影响,但是,那些直接涉及对犹太人看法的回答所受到的影响与研究人员预想的情形一致:访问员的犹太特征越明显,受访者就越有可能对那些带有反犹太主义倾向的回答有所保留(Robinson 和 Rohde,1946)。

种　族

Hyman 等(1954)在 1942 年进行的一项研究关注的是访问员的种族特征对黑人受访者的影响。这项调查向田纳西州孟菲斯城(Memphis)的黑人受访者提出了一系列有关种族状况、族群关系(intergroup relations)以及征战成就(war effort)的问题;结果显示,黑人访问员和白人访问员所获得的回答存在着显著的差异。Hyman 等人对这些差异的解释是,黑人受访者对白人访问员给出的是“他们愿意听到的回答”;而同时在纽约进行的对照实验所显示的差异要小得多。Hyman 认为,这些数据表明在 1950 年代早期,访问员的种族特征对此类问题所产生的影响在南部地区表现明显,而在北部地区则比较微弱;原因可能是北部地区的黑人对

访问员的种族偏见较小，也可能因为他们更愿意对白人访问员表达自己的意见。

在 1960 年代，Schuman 和 Converse（1971）对同样的问题进行了研究。在底特律地区的研究课程即一门给社会学系研究生开设的调查训练课程中，学生们按照肤色进行了分组，分别对底特律地区黑人和白人对照样本的受访者进行访问。这项研究的优势在于其所使用的黑人和白人访问员对访问的了解和熟练程度都十分相似，其缺点则是两组访问员都未受过额外的训练，也没有任何访问经验。

这项研究中 90% 以上的内容没有因为访问员的肤色问题而受到明显的影响。即使是诸如混合制学校的重要性这样一些社会普遍关注的种族问题，也没有受到访问员肤色特征的影响。但是，当问题直接询问对某一肤色群体的态度或观点时，访问员肤色特征对结果的影响就显现出来了。无论是黑人还是白人受访者，都不太会表达出对访问员所属肤色群体的批评和负面意见。

此外，有些回答中存在的差异并不是出于上述原因，而是因为受访者一开始在认知上受到了影响。例如，在差异较大的问题中有一个是询问受访者最喜欢的艺人是谁，而黑人受访者在接受黑人访问员的访问时更多地回答是黑人艺人。这一现象也许可以解释为受访者会按照访问员的喜好来选择回答，但此书作者同时还认为，当黑人受访者在面对一名黑人访问员时会更多地联想到黑人艺人并从中作出选择。这属于情境效应的一种。

Anderson 等人（1988a，1988b）最近的一个研究比较了黑人受访者在投票问题上对不同肤色访问员的不同回答，他们还对照投票记录分析了这些回答的准确性。Anderson 发现，访问员的种族特征的确对回答产生了影响：黑人受访者在选举之前的调查中对黑人访问员表达投票意愿的可能性更大；而在选后调查中，面对黑人访问员他们更有可能回答说自己的确参与了投票。比对投票记录后，研究人员发现，在选后调查中白人访问员所获得的回答更加准确；受访者在面对黑人访问员时会言过其实。在这项调查中，访问员的肤色特征的确对访问情境有显著的影响。而进一步的证据表明，在选举前接受黑人访问员调查的人在选举时参与投票的可能性会显著提高；访问使他们的行为发生了改变。

在 Shuman 和 Converse 的研究中,我们无法看出什么样的访问员所获得的回答最好、最真实。对主观现象的测量无法对回答的真实与否作出判断。唯一明确的是,在一个标准化测量的过程中所得到的回答不应与访问员的某些特质联系起来。如果根据访问员的肤色特征就能预测回答结果,那就说明这一测量在不同肤色的访问员中没有实现标准化。当访问员的肤色特征确实会对回答产生显著影响时,我们可以通过使用相同肤色的访问员来减少测量误差。

访问员—受访者关系

访问员的个体特征已被证明会对问答双方间的关系产生一些可以预知的影响:相互间的好感以及双方关注的重心是人际关系还是调查任务。这些影响与数据质量之间的关系则不甚明了。

社会地位与教育

Fowler(1966)研究了受访者的受教育程度如何影响到问答双方相互间的印象以及双方的交流方式。社会心理学文献中也有很多证据阐述了受访者与访问员之间的相对地位如何影响双方间关系。概括地说,人们更愿意与和自己同等或者比自己层次高的人交流。当存在层次差异的时候,低层次一方为了与高层次一方建立起关系往往会分散对于调查任务本身的注意力;而高层次的一方在与层次较低的一方建立关系时则更关注任务本身(如 Cohen,1985)。我们可以想见,在层次地位等各方面都相似的访问员和受访者之间更容易建立起一个良性的访问关系(如 Newcomb,1961 和 Lundberg 等,1949)。

Fowler 的研究发现基本上与此一致。从受教育水平所反映出的社会层次来看,访问员之间是大体相同的,大多数人都有高中以上学历,有的上过大学。表 6.1 归纳了一些相关结论。总的来说,我们可以看出,访问员事后对受访者的评价与后者的受教育水平直接相关;受访者的受教育水平越高,访问员对他们的印象就越好。此外,如果受访者有高中学历或读过大学的话,访问员更有可能在访问结束后留下来同他们聊一会儿天。

表 6.1 访问员和受访者之间人际吸引的有关指标：依据受访者受教育情况分类

人际吸引指标	受访者教育程度			
	0~8 年	1~3 年高中	4 年高中	大学 1 年及以上
访问员对受访者的喜欢超过平均水平*	54%	55%	66%	73%
访问结束后，访问员留下交谈超过 1 分钟*	54%	49%	67%	69%
受访者非常喜欢访问员	35%	46%	44%	42%
受访者希望访问员是“公事公办”而非“登门拜访”*	49%	65%	38%	28%
N	(129)	(89)	(123)	(67)

资料来源：Fowler(1966)，略有修改。

注释：* 高中毕业生稍不同于 1~3 年的高中受教育者，t 检测值 $p<.05$。

从受访者的角度而言，他们对访问员抱有多大好感并没有什么规律可寻。但是，在问到他们更喜欢访问员表现得例行公事还是上门拜访时，受教育程度较低的受访者表示倾向后者；而那些高中毕业、学历仅比访问员稍低一些的人则最偏好“上门拜访”。

基于这些数据，人们可能会得出推论说，在访问员与受访者层次相当的情况下回答效果是最好的。这一看法会鼓励研究人员在对社会—经济层次较低的人进行研究时雇用一些与受访者背景相似的人作为访问员。但是，Weiss(1968)对这一方法表示了怀疑。

Weiss 对纽约享受福利的母亲(welfare mothers)作了一项研究。访问计划包含了多个话题，其中一些可以对照档案记录进行核实。她评估了四项问题回答的准确性，其中两个是关于投票登记，两个是关于孩子在学校的表现。她发现，问答双方之间在年龄、教育及社会经济地位上的相似性并不会增加回答的真实性；各方面最为相似的问—答组合所得出的数据不见得更加真实准确。相反，有几对组合中，访问员的受教育水平与社会经济地位要明显高于受访者，但获得的回答质量反倒更高。

Weiss 在研究中还要求访问员对自己对受访者所表现出的亲和感作出评价。奇怪的是,“融洽”的程度越高,结果数据的偏差也越大。Weiss 的研究中有两点最重要、也最具说服力:其一,个体特征上的相似性并不能改善测量的质量;其二,亲和感与相互间的好感对调查而言并不一定是好事,反而有可能会对数据的质量产生负面的影响。

总之,已有的研究文献没有提供任何应当根据受教育水平或社会地位来选择访问员的依据。这其中的关系是复杂多变的,对不同受访者发生作用的情况也各不相同。最为确定的一点是,我们完全没有理由仅仅为追求问答双方的相似性就选择一些相对缺乏经验的人,而舍弃那些资质更深的访问员。

访问员的年龄

将访问员的年龄作为调查的一个影响因素而进行系统分析的研究并不多见。Erlich 和 Reisman(1961)研究了访问员的年龄对青少年回答的影响。结果与先前的假设一致,即年长的访问员所获得的回答比较正式,更像是父母愿意听到的话。Erlich 和 Reisman 认为,20 多岁的访问员所得到的回答要比 30、40 岁的访问员所得到的回答更能反映真实的情况。

年龄起作用的方式与社会地位相类似:年龄差异越小,双方就越容易沟通。较大的年龄差异会产生代沟,使双方的交流变得拘谨,从而有可能使沟通产生障碍。但是,Weiss(1968)的研究并没有发现年龄会带来任何影响。此外,就我们所知,还没有一项研究能证明访问员的年龄会导致误差。

访问员的性别

Hyman 等(1954)指出在下面这个问题上,女性受访者对不同性别的访问员所给出的回答是不相同的:“优秀的男士都不会尊重那些有过婚前性行为的女性。”如果访问员是位男性,她们明显更倾向于对这一说法表示赞同。

我们的研究同样显示出,受访者对不同性别访问员的评价是存在差异的。我们对一项健康调查中的受访者进行回访,征询他们对所受访问的反馈。他们就 10 项不同的指标对访问员进行了

评价。在所有 10 项指标中,无论男性还是女性受访者对女性访问员的评价都要高于男性访问员,其中几项差异更加显著。表 6.2 列出了对亲和性、专业性及总体表现的评价。

表 6.2 受访者对不同性别访问员的评分

受访者评分	访问员性别		
	男性	女性	p
总的来说,访问员的表现“优秀”	44%	62%	<0.01
*访问员的“亲和力”得分为 10	36%	54%	<0.01
*访问员的“专业性”得分为 10	52%	64%	<0.05

注:*从 1 到 10 的量表,10 代表最高。

同其他研究者一样,我们也对这些不同性别的访问员样本的可比性存在怀疑。因为接受临时性兼职工作的男性相对较少。此外,相对于男性而言,访问员这一行对女性的吸引力要更大一些。我们的研究只包括了 9 位男性访问员。我们还要强调的是,这些数据是通过入户访问收集的。电话访问中的性别影响肯定要小得多。

我们相信这一问题领域还需要进行更多的研究。我们反对任何那种超越了具有复合特征的宽泛分类群体所推衍出的通则。而访问员的性别分类显然就属于这种宽泛的分类。此外,上述这些差异并不能证明女性访问员收集的数据质量就更高。这一假定缺乏完善可靠的研究论证。虽然我们的数据足够明晰,但是我们仍然不能确定在大多数调查访问所建立的这种即时关系中,性别究竟是不是一个重要因素。我们希望今后的研究能为此提供更多的数据。

结 论

对于大多数研究而言,选择访问员的依据只有一个,那就是做好这项工作的能力和意愿。访问员特质与访问员相关误差之间还没有发现规律性的联系。但是,当调查的主题直接关系到访问员的某些特征时,受访者可能会认为有些回答项目会对访问员

造成不敬或是冒犯。这种情况下研究人员可能会对这些访问员特征严加控制。但是,即使在目前为止最为深入彻底的研究中(Shuman 和 Converse 对不同肤色的受访者和访问员的研究),也只有不到 10% 的问题在回答时会受到访问员种族特征的显著影响。

有许多种不同的访问员特征和行为都会对问答双方间的关系产生影响,单独列出其中一个是毫无意义的。从访问员人为误差的角度来看,我们的建议总的来说就是,尽可能使用最好、最符合标准化要求的访问员。

但是,在研究了受访者对不同性别访问员的不同回应之后,我们开始思考,关于访问员性别对受访者影响的研究是否足够充分。考虑到我们的发现,我们认为,如果没有同时发现性别对数据的其他影响,那才是令人吃惊的。然而,到目前为止,确实没有这方面的数据。因此,对于那些想要降低访问员人为误差的人而言,需要关注的重点不在于访问员的挑选而在于培训和管理。在接下来的两章中我们将要讨论这两个话题。

第7章

访问员培训

Training Interviewers

事实上，计划进行严格调查的研究者中，没有人会质疑访问员培训的价值。之前各章所提到的标准化访问的各种规则、技巧和程序，都不能指望访问员会自觉自发地加以使用。此外，培训的最佳内容或者必要内容，用于培训的技术以及培训时间的长短，都有很多问题需要探讨。

我们认为，任何特定的调查都需要通过培训来说明其特定的程序和目标。由于数据收集的工作千差万别，因此基本上不可能总结出一个放之四海而皆准的培训方式。一些大型调查计划包含有许多复杂的定义和程序，如全国健康访问调查，因此，原则上需要进行为期数天甚至更长时间的面对面的强化培训，以使访问员完全熟悉访问的流程。但是，目前美国调查组织流行的做法，是将书面材料邮寄给训练有素且经验丰富的访问员，通过这种方式向他们传达各个调查项目的流程以及具体要求。

但是，这里我们所关注的是对访问员就标准化访问技巧而进行的起步培训，我们称之为基本的访问员培训。不论调查项目对访问员有何特殊要求，对访问员进行一般访问技巧和流程的培训总是有必要的。

有关培训的若干选项

访问员的工作通常可以分为四部分：

1. 联系受访者，争取合作；

2. 与受访者建立关系;
3. 处理问—答过程;
4. 记录答案。

一般来说,所有的访问培训都会或多或少地涉及这四项内容。

督导可以使用多种方式培训访问员。他们可以使用:

1. 访问员手册;
2. 讲座;
3. 示范;
4. 指导练习(supervised practice);
5. 陪同演练(monitoring performance),并在培训结束后给予评估和反馈。

在正式访问之前进行的所有指导性活动都可以算作是培训。严格说来,在数据收集完成之后进行的监测和指导则都属于督导(supervision)的范围。而督导将在第8章着重进行讨论。

培训时间的长短显然密切关系到培训所包含的内容、技巧,以及培训的成本。在学术和政府性质的调查组织中,基础性的访问员培训通常持续2~5天;电话调查访问员的培训时间要比面对面调查的访问员更短。但是,有很多调查所使用的访问员,其所接受的基本培训还不满一天。

访问员所接受的培训的差异至少说明三个问题:

第一,有些研究者显然对访问员表现优劣所带来的后果缺乏重视。这种想法可能是因为他们对一些针对访问员影响的研究缺乏了解,也可能是因为他们的研究目标可以容许很大的测量误差。

第二,在"培训进行到什么程度才足够或有助于提高数据质量"这一问题上,专家们还未达成共识。当然,有些研究者认为,对于那些具备一定基础的访问员而言,数小时的培训已足够将他们的表现提高到一个可接受的水平上了。

第三,访问员所接受的培训量与他们的工作条件之间有着明显的关联。对电话访问员的培训时间一般较短,因为培训方认为他们可以对访问进行密切的监测,可以及时发现问题并对访问员

进行必要的纠正。相反,对于面对面的访问则通常无法进行这样密切的监测,因此,督导在访问正式开始之前会尽可能确保访问员作好足够充分的准备。

本章的大部分内容来源于我们对培训和督导是如何影响访问员的一些研究(Fowler 和 Mangione,1986)。其中,我们将特别考察培训对如下方面的影响:

1. 访问员的工作态度(orientation);
2. 访问员的技巧;
3. 访问员与受访者沟通的方式以及受访者的回应方式;
4. 所收集数据质量。

关于访问员培训价值的一项研究

我们雇用了 57 名之前没有任何相关经验的访问员,其中 48 名为女性。我们将其系统地分为 4 组,给每组提供的培训时间和方式各不相同。对其中两组的培训项目仿照学术调查组织通常使用的培训流程,第一组为期 2 天,第二组为期 5 天;对第三组的培训方式则是尽可能地缩短培训时间,访问员仅仅接受了半天的培训就被派出去进行调查;最后一组为期 10 天,大大高于传统标准,用以证明过量的培训是否有助于效果的提升。

其中,不满一天的培训项目包括有 2 小时的讲座,介绍访问技巧,并进行一个示范访问。

为期 2 天的培训,除了以上两项内容外,还放映了一部关于访问技巧的影片,就访问流程进行了讨论和提问,并进行了一定的指导练习。

为期 5 天的培训,与为期 2 天的培训项目相类似,但各个部分都进行了扩充:访问技巧的讲解时间更长,有更多的机会可以提问和讨论;指导练习(supervised practice)的机会和内容也更多,不仅有访问练习,还包括各项访问技巧的练习,如对研究项目的介绍、提问、追问以及对回答的记录。

为期 10 天的培训项目包含有为期 5 天培训的所有内容,此外还加入了三种教学方式。第一,我们训练访问员像督导员(supervi-

sor)那样,依据研究中心的常规基础对访问技巧进行评估。通过这种方式,访问员能够有更高的自觉意识,对高质量访问的标准有更深的认识。第二,访问员阅读一些有关调查中访问员相关误差的文章;每篇文章的读者向小组提交阅读心得,在小组进行讨论。第三,每位访问员进行一次入户访问的练习,督导员进行观察。

如前文所述,培训的内容和技巧与其他各组相同,只是练习的程度以及学员对基本原则的理解程度不相同;此外,学员的实践练习也比其他组别更多一些。

表7.1总结了各培训项目的内容。时间最短的培训项目和为期两天的培训项目的主要差别在于讲座的长度不同,关于访问程序和具体研究程序的讨论的时间也不同。在为期两天的培训中,有105分钟的指导练习,时间最短的一组只是旁听了一次示范访问,没有进行任何指导练习。为期5天的培训项目在为期2天的培训项目的基础上,关于讲座的讨论时间额外增加了1个小时,指导练习的时间则增加了7个小时。

表7.1 不同培训时间的培训工作总结(以分钟为单位)

	培训项目的时间			
培训内容	<1天	2天	5天	10天
一般管理(General Administration)				
关于报酬表、保密规定及一般工作程序	50	75	135	135
一般的访问程序				
电影	0	90	90	90
讲座/讨论	120	180	420	420
示范	55	15	25	25
指导练习及演练	0	105	455	575
*其他访问员报告/讨论/实践	0	0	0	1 140
研究概要				
关于某些问题的目的及研究程序的介绍	75	165	180	180
总时间(分钟)	300	630	1 305	2 565
(小时)	5	10.5	21.75	42.5

注:培训开始之前,所有访问员都被要求阅读手册,并做一些与填写报酬表(pay forms)及选择受访者的程序有关的练习。时间估计为1.5小时。

*实践包括阅读与调查误差来源有关的论文并报告给整个小组;学习使用标准的督导表格对访问员绩效进行评估;就如何编码进行训练,并且就规划和组织调查行程进行练习。

在培训完成后，我们分给访问员40个地址，让他们去做一个半小时的健康访问。关于访问员的表现，有五个不同的评估来源：

1. 在培训的最后，每位访问员完成一次访问练习，我们进行录音和评估。
2. 在每次健康访问结束之后，由另一名访问员对同一名受访者进行一次后续的电话访问，了解其对此次健康访问及访问员的反馈情况。
3. 我们要求访问员在完成任务后填写一份问卷。问题包括先前的访问，访问员对此次工作的感受，以及一些测量访问员技能的测试题。
4. 我们让三分之一的访问员对他们的访问进行了录音。对这些访问录音进行编码能够提供更多有关这部分访问员的信息。
5. 最后，由于每个访问员所分配到的地址都是从整个样本中随机抽取的，因此有可能对这些访问所获得的回答进行分析，并评估这些访问员对各自所获回答的影响程度。

培训对访问员工作态度的影响

我们从访问员对于培训经验的感受谈起。作为一个信息来源，培训的重要性有多大？我们发现，访问员接受的培训时间越长，访问员就会更多地将培训作为一个重要的信息来源（而不是访问员手册，相关工作经验，或者督导的反馈），这并不让人吃惊。当我们要求访问员在“很好”到“糟糕”的量表上对培训作出评价时，结果符合预期，受过10天培训的人显然评价更高。不过培训时间最短的两组没有表现出明显的差异（参见表7.2）。

需要指出的还有，培训时间最短的两组成员将指导手册列为最重要的信息来源。在解释本章的这些发现时，我们需要牢记的是，培训时间最短的这一组成员都持有很好的指导手册，并且都处于良好的督导之下。虽然他们的培训时间很短，但是相比于那些培训时间同样很短但却没有好的指导手册和督导的访问员，他

们对于如何正确进行访问的知识要懂得多得多。

表7.2 访问员认为在与访问技能相关领域的准备工作中"低于需要"者所占比率:依据培训项目的时间进行分组

	培训项目的时间				
技能	<1天	2天	5天	10天	p*
a. 计划行程并有效利用时间	30%	46%	23%	7%	.02
b. 向受访者解释研究目的	13%	15%	21%	33%	n. s.
c. 争取受访者合作	33%	77%	43%	20%	.02
d. 提问	20%	8%	0%	0%	.04
e. 非诱导性追问	54%	70%	14%	7%	.02
f. 实现提问目标	33%	62%	36%	20%	.14
g. 记录答案	27%	15%	14%	7%	n. s.
h. 处理访问中的人际关系	33%	8%	14%	0%	.10
N	15	13	14	15	

注:计算概率时,先将训练时间为1天和2天的合并为一组,将培训时间为5天和10天的合并为一组,使用2×2表计算卡方值。

另一个问题是,访问员自认为准备得如何。注意,这些评价都是在访问员完成访问以后做出的。表7.2列出了这些结果,并为我们提供了许多有趣的信息。

总的来说,我们可以看出,访问员对自己准备程度的评价与他们所接受的培训量是相互关联的。这些关联并不都具有统计学上的显著性:只有3项是完全线性的;但是除了2项之外,其他都与预期的情形大致相符。

与预期不符的2项分别是介绍研究和获得合作。在这两项中,培训时间最短的访问员对自己准备程度的评价与其他人一样。此外有一点很有意思,在寻求合作这一项上,结果显示他们与其他各组成员做得同样成功。

从问—答过程及访问的处理上来看,也可以发现接受5天培训的一组和接受10天培训的一组在各个维度上都有一些差异。但是,在"非引导性追问"和"实现提问目标"这两项上,接受2天培训的一组对自己准备程度的评价要大大低于培训时间更长的两组,甚

至比培训时间更短的那一组还低。在表 7.2 所示的 8 项评分中的 4 项评分,其接受 2 天培训者对自己的评分要比其他各组都低。虽然这些数值都很小,而且单个偏差可以被视为统计错误,但是这一情况的反复出现必定有其含义,对此我们将另作评论。

最后,表 7.2 显示出,在有些方面,培训时间最短的组员最有可能认为自己准备不足。显然,最棘手的一项是如何进行追问,这点在我们看来是标准化访问中最为关键的访问技能。在培训时间少于 5 天的组员中,大部分人认为自己在非引导性追问方面接受的培训不足;很大一部分访问员认为,对于表中的其他 3 个方面——计划行程及有效利用时间,寻求受访者的合作,以及实现提问目标,他们可以准备得更好。这些数据很好地为我们指出了在培训中哪些方面最需要给予特别的重视。

培训对访问员的第三个影响是他们对工作中不同方面的优先性排序。培训时间最短的一组与其他各组相比,最关键的区别是对问答技巧的培训时间不同,尤其是提问、追问和记录。因此,我们有理由认为,他们对标准化的重视程度要低于其他人。通过列表比对,我们可以发现,各组成员都将“收集精确、完整的回答”列为首要目标。但是,与我们预想的一样,培训时间较长的组别与培训不到一天的一组相比,对标准化的重视程度要更高。

最后,访问员被要求回答自己作为一名访问员的感想如何。回答结果与培训量没有直接关联,但是,培训期为 2 天的那一组再次表现得与众不同,他们最不乐意当这一次的访问员。

访问技术

访问员培训最直接的目的当然是教会他们如何介绍研究项目,引导受访者,以及处理问答过程。在我们的研究中,我们就访问员如何介绍访问项目收集了两类信息。首先,在事后工作报告中,我们要求访问员回答一系列受访者在开门后可能会问的问题,并对其回答的内容和准确性进行了编码。总的来说,这些卷面回答的质量与培训时间的长短无甚关联。

第二类信息来自对受访者的回访。我们询问受访者他们对此研究项目了解多少,以及对于他们的提问访问员回答得如何。

对于后者,受访者的评价与访问员所受的培训没有关系。但是,受访者对研究目的的了解程度与访问员的培训时间成反比:培训时间为1天的一组所访问的对象对研究目的的了解略高于平均水平,而培训时间为10天的一组要明显低于平均水平。卷面测试以及培训过程本身清楚地说明这一结果不是因为10天组的培训缺少了某些信息。我们认为,原因在于对访问员的培训方式的差异。

在培训结束后随即进行的访问练习的录音为我们提供了更多信息。虽然由于第一组仅有15名组员,因而所有这些人际关系练习并不具备统计学的意义。但是,表7.3显示,培训量最少的一组在提问和追问的表现上不尽如人意。表格还显示出其他三组在访问技能上并没有很大的差别。

表7.3 依据访问录音对访问员行为进行评估:按照培训项目的时间进行分组

"优秀"和"满意"访问员所占百分比	培训项目的时间			
	<1天	2天	5天	10天
逐字逐句读问题	40%	62%	79%	60%
追问封闭性问题	53%	69%	79%	73%
追问开放性问题	13%	31%	29%	47%
记录封闭性问题的答案	87%	85%	100%	80%
记录开放性问题的答案	67%	69%	71%	60%
人际行为无偏误	80%	92%	92%	93%
N	15	13	14	15

注:本表所有这些相关性都没有达到一般的标准,总体上不具有统计显著性。

我们之前已经提到,每组有三分之一的组员要对自己的访问进行录音,除非受访者拒绝(这种情形很少见)或是录音设备出现故障(相对更常见)。因此,对于这些访问员,我们能够详细地了解他们是如何操作问答流程的。表7.4汇总了这些数据。

表 7.4 依据录音编码对访问员行为进行评估：按照培训项目的时间进行分组（仅分析进行第三个层次督导的访问员）

	培训项目的时间				
	<1 天	2 天	5 天	10 天	p**
录音编码反映的访问员行为					
没有读错问题/访问出错的总次数	21	7	14	6	<.01
诱导式追问/访问的总次数	8	5	5	3	<.01
没有进行充分追问/访问的总次数	8	6	5	5	<.01
没有正确记录封闭式问题/访问的总次数	1	1	1	*	<.05
没有正确记录开放式问题/访问的总次数	4	2	2	2	<.01
不合适反馈/访问的例子	2	*	*	*	<.01
优秀或者满意的访问所占百分比					
逐字逐句读问题	30%	83%	72%	84%	<.01
追问封闭性问题	48%	67%	72%	80%	<.01
追问开放性问题	16%	44%	52%	69%	<.01
记录封闭性问题的答案	88%	88%	89%	93%	n. s.
记录开放性问题的答案	55%	80%	67%	83%	<.01
人际行为无偏误	66%	95%	85%	90%	<.01

注：* 少于 0.5 次。

** 依据 *F* 检验，自由度依据访问员得分的组内相关系数（intraclass correlation）调整。

分析依据是 20 位访问员的 320 次访问，平均每位访问员进行 16 次访问。

对于这些作了录音的访问员，结果是很明晰的。培训量少于 1 天的学员在访问时的表现普遍不尽如人意。除了封闭型回答的记录之外，这一组学员几乎在其他各方面都明显差于其他各组。此外，我们还能看出，在这些表格所列出的绝大多数技能中，培训时间为 2 天和 10 天的两个小组都没有显著的差异。但是，在追问这一项上，访问员接受培训的程度与他们的实际表现之间存在着稳定的关联。数据再次显示，追问是问答过程中难度最大的一个环节，对访问员的培训越多，越能够提高他们在这方面的能力。

受访者对访问员的看法

为什么我们说培训会影响到访问员与受访者建立起的关系呢? 首先,培训使得访问员在工作时更加得心应手,对访问的处理也更为自信,这就会给受访者留下一个更加专业、干练(competent)的印象;其次,访问员在培训中得到明确的要点就是,强调要以访问任务本身为导向(task orientation)。我们不鼓励访问员与受访者建立起私人性质的联系,因为这样很可能导致偏差和其他一些我们不乐见的结果。而通过更多的指导练习和针对访问员作用的扩展讨论,访问员对这些原则的认识会更加明晰。

在验证这些假设的过程中,我们遇到了这样一个问题,几乎所有的受访者对访问员的评价都很积极(positive)。例如,当我们要求受访者在一个从 1 到 10 的量表上对访问员的专业性和亲和性进行评分时,平均分都超过了 9。这一发现可能说明,受访者普遍不愿对访问员提出批评,同时也更愿意对访问员作出真正积极的评价。

不管出于什么原因,总之我们几乎没有在受访者的反馈中找到任何证据表明,在人际交流方面,培训程度不同的访问员之间存在什么差异。例如,在亲和性、专业性以及访问员是否保持中立且没有发表个人观点这些方面,受访者对不同组别的访问员的评价都相同。我们还询问受访者他们如何定性各自的访问员,是"公事公办型"还是"上门拜访型"。其结果也与培训项目无关。

然而,倒是有一些差异表明,培训时间长的访问员更加以访问任务本身为导向,而培训时间短的访问员则更侧重于人际交流。在访问任务方面,我们询问了受访者他们的访问员对数据精确度的要求有多高,他们是被要求提供确切回答还是"大致想法"即可;这些问题意在测量访问员向受访者提出要求及指导其回答的意愿。在"访问员是否要求确切回答"这一项上,受访者的评价存在着一个具有统计学意义的规律,即培训时间少于一天的小组向受访者提出这一要求的概率最低(见表 7.5)。此外,访问员对精确度的重视程度也表现出相通的规律。

表 7.5 受访者的评分：按照培训项目的时间进行分组

受访者认同以下表述者所占百分比	培训项目的时间				
	<1 天	2 天	5 天	10 天	P*
访问员要求“精确”的答案（而非一般的想法）	70%	78%	82%	77%	.01
访问员工作很“优秀”（参照为很好、好、一般和差）	64%	57%	57%	55%	.02
N	376	310	357	342	

注：* 卡方检验，将其他各项合并，与接受培训小于 1 天者进行比较。

在人际交流方面，对亲和性及是否表现得公事公办的评价没有显示出预计中的关联。但是，培训时间最短的一组一定在某些方面表现得很好，因为在总体表现这一项上，受访者对他们的评价最高。同样，这一组受访者对访问的兴趣也最大。

这些差异并不很大，而且我们之前已经说过，受访者在评价访问员行为的时候不是很准确客观的。有些情况，例如访问员是否表现得客观中立，受访者的评价与我们对访问录音编码分析的结果并不吻合。但是，我们还是认为，的确存在一些证据能够证明，培训能够加强访问员对访问任务的侧重，而减少其对人际关系的关注。

培训对数据质量的影响

在这一研究中，我们通过两种方法测量数据质量：

一是 rho，即组内相关系数（intraclass correlation），用以测量访问员对其所获回答的影响程度；其效果也就是测量一组访问员的标准化程度。组内相关系数的理想值应接近零。

二是对于这次调查所选取的 54 个问题，我们能够预估出质量更高、更精确的回答会偏向哪一边。我们的预计是基于一个“社会期望假说（social desirability hypothesis）”，即社会普遍认同的回答越少，偏差就越小。作出预估的另一个根据是“漏答假说”（underreporting hypothesis），即细枝末节及不易回忆的事件一般在回

答时会有所遗漏。因此,我们将每个访问员所得到的回答与整个样本的平均值进行比较,并计算出培训时间的长短对访问员所获回答的质量的影响。

这两种测量方法都没有发现任何具有统计学意义的关联。对数据偏差的测量显示,培训有时能够提高回答的质量,培训时间最长的一组所收集的数据偏差最小;但是,培训时间为半天和5天的两个小组并没有任何差异,因此很难得出一个确切的结论。

在标准化方面,数据则更加复杂。标准化程度最高的数据均出自培训时间为2天或5天的访问员。虽然样本的规模并不能产生极具说服力的统计数字,但似乎培训时间最长和最短的两组都比中间两组表现得更差。毫无疑问,长达10天的培训也没能提高这些访问员的标准化水平。

结　论

一个明确的结论是,访问员需要在一般访问技巧方面进行指导练习才能够成为一名合格的访问员。仅仅使用阅读、讲座及示范的方式进行培训是不够的。接受最低程度培训的访问员也能够像其他人那样处理好诸如寻求合作、与受访者建立联系等人际交流问题,尽管这在一定程度上是因为我们不是很了解如何更好地对访问员进行这方面的培训。然而,他们对于问答环节都感到准备不足,并且在录音记录中,大部分组员的表现都不尽如人意。此外,虽然数据质量分析的结果也不是十分明晰,但还是有证据表明,在数据偏差和标准化这两个方面,他们所得的数据质量显然不如其他各组成员。Billiet 和 Loosveldt(1988)在比利时进行的一项研究将完全未受训练的一组访问员与受过3天培训的另一组作了比较。他们同样发现,在需要进行追问的问题上,访问员的处理方式存在着明显的差异。显然,追问技巧对标准化访问而言至关重要,也是最难学的技巧,而更多的培训有助于访问员对这些技巧的掌握。

接受为期2天培训的一组访问员显然掌握了基本的访问技巧。他们的培训计划包括了更多的示范、讨论,以及一些指导练习。对访谈录音的编码分析显示,培训时间更长、尤其是接受了

一整天指导练习的组员在开放性问题的追问上要表现得稍好。但是,从数据质量的角度而言,这组学员所得数据的标准化程度及偏差度与其他各组均无差别。这组学员的欠缺之处在于,与其他学员相比,他们自认为准备不足,而且不如其他学员那么乐意去进行访问。但如果访问是通过电话而不是入户访问的形式,情况则不会如此。但是,我们的数据显示,如果这些访问员能接受更多的培训,他们的信心和兴趣都会得到提高。

而为期 5 天,尤其是 10 天培训的两组数据表明,过多的培训反而适得其反。在设计 10 天培训计划时,我们假定访问员对正确的处理方式了解得越多,他们在实际访问中作出正确处理的可能性就越大。我们没有意识到过多的培训会让访问员感到厌倦。在为期 10 天的培训结束之后,这组学员在访问练习中的得分比培训期为 2 天和 5 天的两组学员还要稍低一些。最重要的是,与后两组学员相比,他们所得数据的标准化程度也要更低,其差异接近统计显著性。Cannell 等(1977a)就为我们提供了几个访问员越学越糟的例子。随着培训量的增加,他们非但没有进步,反而热情耗尽,变得懒散、粗心。虽然我们的结论缺乏权威性,但显然也说明了过长的培训时间导致了相同的情况。

基于上述数据,我们得出一个毋庸置疑的结论,即访问员的培训必须包含有指导练习。它能有效提高访问员的技能和数据的质量。低成本的录像设备能够使这种培训方式获得额外的效果。我们可以对访问员的指导练习进行录像,并在小组讨论时播放这些录像带,这种方法能很好地帮助大家讨论和提高。

访问培训的最佳长度取决于培训小组的规模以及他们所要参与的调查项目的复杂性,还可能与其在培训结束后所受的监督方式有关。但是平均而言,一次培训的合理标准是 2 ~ 4 天的基本技能培训,辅以大量的指导练习,这对于大多数访问而言是最优的选择。但是,培训的效果与培训结束后访问员所受到的监督也有关系。下面我们将在第 8 章讨论这一问题。

第8章 访问员督导

Interviewer Supervision

督导是指收集信息以了解访问员在数据采集之时的表现，对这些信息进行评估，并将评估反馈给访问员的过程。我们将培训(training)和督导(supervision)进行区分，前者是在访问员开始采集数据之前展开，而后者则是在访问员开始采集数据之后实施。

调查机构可以从如下五个不同方面对访问员进行督导：

1. 他们完成了多少工作，这可能仅仅指某一时间段内访问的数量，也有可能包括其他工作；
2. 他们花了多长时间；
3. 他们的回复率，即他们访问成功的人数占分派给他们的受访者人数的百分比；
4. 他们完成访问的质量；
5. 他们处理访问过程的方式，以及与受访者互动的方式。

用于督导的信息

不同的调查机构在对其访问员的监管方式上有很大的差异。事实上，所有的调查机构都会对访问员的工作时间和工作量进行记录。但是，要衡量一次访问的效率和效益，通常不会像计算每次访问的耗时和耗资这样简单。这种简单换算的前提是不同的访问任务之间具有可比性。但是，有些访问很难找到受访者；另外，有些电话访问如同大海捞针，成功率极低；这种因素所带来的成本提升，与访问员的表现是无关的。此外，在一项研究的进行

过程中，也很难对其效率作出合理的评估。但是，几乎所有的调查机构都会计算访问员每次访问所耗费的成本，并以此作为考核访问员表现的依据之一。当然，成本对于任何一个调查机构而言都是十分重要的，但是，它们与所采集到的数据质量几乎没有任何关联。

还有一个常用的方法则是计算访问员的访问成功率。假如我们分配给每个访问员一定数量的受访者，由他们负责进行访问，那么我们就很容易计算出这个访问员的访问成功率是多少。同样，访问成功率的计算要将访问本身的特性计算在内，如果某些访问员所分配到的样本本身就比其他人的受访者更容易联系上，或表现得更加配合，那么就不能以此来断定这个访问员的表现更为突出。

此外，在一项研究的进行过程中对受访率进行检测也是很困难的。在访问员的访问任务大致完成之前，对他的访问表现和成功率是不易进行评判的。有一种方法是让访问员在调查的进行过程中向督导人员汇报有多少人直接拒绝了访问，有多少人对访问持排斥态度。这种方法固然有一定的帮助，但其效果取决于访问员的汇报质量，而这一点往往是很难确保的。

如果样本任务是分配给一个访问小组，而不是具体到某一个访问员个人，那么对访问成功率和访问员效率的考核就更为困难了。在电话调查中，这种将样本分派到小组的做法是最为常见的，因此对个体访问员的评估也就尤为困难。虽然监督人员可以分别计算某一访问员访问成功和失败的次数，但是却很难判断这些访问员所接触的样本在小组中是否具有代表性。此外，一次访问会不会被界定为“拒访（refusal）”，在一定程度上取决于访问员自己对这次访问的描述。同样一种被拒绝的情形，有些访问员会界定为“失败”，而有些访问员则会将其归因于受访者当时很忙。

此外，大多数调查机构会对访问员所收集的信息是否全面充分以及访问员对访问守则的遵守情况进行评估。这种评估是通过对已完成的访问进行复核来实现的。大多数（但不是全部）调查机构有一系列的规程，用以对访问员在数据采集过程中的规范程度进行系统的复核，包括访问员是否遵照规范流程，答案的记录是否恰当，所得回答是否符合要求等。不同的调查组织在复核

的内容以及处理问题的方式上存在着很大的不同。

通常最常见的做法是对访问员所完成的第一个或前两个访问进行复核,如果发现严重的问题则告知并提醒访问员。另一种较为彻底的方法是在所有的访问对象中选取一个样本,随后对这些样本的访问进行复核审查,并且定期向访问员提供书面或口头的反馈;这样做无疑会有助于提高这项调查的整体质量。如果不作定期反馈,那么至少在调查结束后,要将审核结果反馈给访问员,这样也会有助于他们在今后的访问中避免同类的问题。最彻底的做法是对访问员所进行的每一次访谈都进行复核,从而保证每次访问都能采集到预期的数据。如果一次访问未能采集到调查所需的全部有效数据,有些调查组织会要求他们进行回访。对于一份很长的问卷,我们很少会将所有的问题进行回访;但针对其中一些关键问题,这种方式还是较为常见的。

本书的写作目的在于确保调查研究中测量的质量,就这个目的而言,上述这些监督方法都不是关键所在。这些监督方式很少会影响到访问的质量,影响到访问员对访问的完成状况。事实上,监督人员只有通过特殊的方法才能对访问过程中的数据采集情况进行监督。仅仅通过观察一次已经完成的访问结果,我们很难判断出相关数据是如何采集到的。如果开放性问题被认为是逐字逐句记录的,人们可以仔细阅读答案,从而很容易识别出哪些是进行了概括或者解释的。很多调查机构要求访问员对自己所用的追问方式进行记录,如果记录规范,监督人员就很容易看出这些追问是否适当。显然,这种复核对访问员的要求不高,但要成为一名优秀的标准化访问员,仅仅做到这些还是远远不够的。

如果不对整个访问过程进行全程监督,我们就无从知晓访问员是否严格按照问卷书面的表述进行提问,无法对追问的质量进行监督和评估,不知道访问员所使用的追问方式是否具有诱导性。监督人员无从判断访问员在记录答案时是否审慎,也无法对问答双方之间所建立的关系作出任何评估。

如果不能对访问员的访问方式进行监督,将会导致两个后果。第一,当然无法鉴别访问中存在的问题,不能筛选淘汰那些不能很好完成访问任务的访问员;第二,无法使访问员意识到标准化访问的重要性。如果一个调查研究组织花费数天时间培训

访问员如何进行标准化访问，却在调查进行过程中没有对访问的质量和标准化程度进行监督，那么访问员就很难在调查过程中始终将标准化作为首要目标。此外，如前所述，调查过程中的许多因素，比如受访者更喜欢轻松随意的交流氛围，访问员想要高效快速地完成访问，这些都会对访问的标准化产生负面的影响。因此，从管理的角度而言，似乎有必要对访问的问答流程进行常规的监督。

对访问质量进行监督的方法步骤因访问地点的不同而有所差异，电话调查和入户调查所用的监管方法也是不一样的。

电话访问的监听

大多数监控电话访问设备配备有专门的监听电话，督导人员可以在各个电话访问中进行切换而不被发现，因而不必打断访问过程。所有的电话访问设备都应当配备这样的功能。

通过随机监听电话访问，监督人员得以辨别出哪些访问员工作不够称职。但是我们认为，高效的监管需要更为系统的方法。我们认为，对电话访问的有效监管应当具备以下四个要点：

1. 要对每个访问员进行定期、频繁的完整电话监听。我们所说的频繁，是指在至少有十分之一的电话访问要得到完整的监听。
2. 监听电话的监督人员应当受过专门的培训，懂得如何评估访问员的表现和访问的质量。有一个要点必须注意，那就是当监听人员在一人以上的时候，必须通过系统培训来统一他们评判的标准。
3. 监听人员应当填写一张系统评估的表格，对所监听的每次访问进行记录。从我们的经验来看，如果监督人员只是简单地监听和记录错误，那么就很容易出现评判标准不一致的情况，而且难以对访问的各个方面进行全面的关注。而在这种情况下，他们更容易关注访问中存在的问题，而忽视访问员表现突出的方面。表 8.1 给出的就是这样一张监听表格。Cannell 和 Oksenberg(1988)也给出了一套类似的步骤。

表 8.1 监测表格(Monitoring Form)

姓名 督导 从介绍或者问题_____ 到问题_____ 反馈日期_____ 评分______		
A)介绍	确认电话号码【 】	而且是住宅电话【 】
B)申明	姓名【 】 资助方【 】	目的【 】
C)解释	自信【 】 主动【 】	可以跳至问题【 】
1	按照书面表述准确读出问题 不正确提问________________	20【 】完全正确 15【 】1~2 个不正确 10【 】3~5 个不正确 0【 】超过 5 个
2	追问不具有诱导性，并且合适 诱导性追问或者不合适的追问(列出问题和追问)________________	20【 】完全正确 15【 】1~2 个不正确 10【 】3~5 个不正确 0【 】超过 5 个
3	有必要追问之时，没有追问 问题编号________________	20【 】从未 15【 】1 次 10【 】2 次或者 3 次 0【 】超过 4 次
4	跳答不正确 问题编号________________	20【 】从未 10【 】1 次或者 2 次 0【 】3 次或者更多
5	反馈：对答案进行个人性的或者不合适的反馈 问题/评论________________	20【 】从未 10【 】1~2 次 0【 】3 次或者更多
6	指导：必要时，解释研究的作用、目标和原因评价________________	20【 】是的，需指导/很好 10【 】是的，但可以更好 20【 】没有需要指导 0【 】没有，或者做得很差
7	节奏	20【 】很慢 10【 】中等 0【 】太快

4. 督导人员在所监听的访谈结束后应当立即与访问员沟通，讨论表格中所记录的内容。这种及时的反馈是最有效的，它能够确保访问员对自己在访问中的优劣表现都有一个明确的认识。

面对面访谈的监督

如果访问不是在某一个地点内集中进行的，那么对问一答流程的监管显然要困难得多。很多年来，一个标准的做法是派一名监督员陪同访问员进行访问，对培训结束后的第一到二个访问进行监督。偶尔也会对之后的访问进行监督。这种方法能够看出访问员是否能够胜任访问工作，但却不容易看出访问员在长时间内是如何进行访问的。此外，考虑到时间和成本因素，这种监督在最初的培训结束之后是非常少见的。资历较深、经验丰富的访问员一般不会受到这样的监督。

廉价便携式录音机的问世使得对面对面访谈进行常规监督成为可能。监督人员可以要求访问员对其所有的访问都进行录音，或者选取一部分样本进行录音；然后，根据录音进行评估。当然，录音需要得到受访者的同意，但一般来说这都不是问题。访问员只要向受访者解释清楚录音是为了确保访谈的质量，一般都会得到理解，不会遭到拒绝。

对于通过录音进行的监督，我们有以下几点实用性的建议：

1. 应当要求访问员对某一时段内的所有访问都进行录音，而不只是择其一二。相对于不同的受访者，对其访问的难度也是各不相同的。遇到较“好”的受访者，访问员的表现自然也会更好。因此作为一名监督员，应当考察同一个访问员面对各种不同受访者时的表现。
2. 假设所有的访问都有录音，那么监督员则可以决定选择听取其中一定数量的录音。当然我们有各种各样的理由认为只要从所有的访问中选取一定的样本进行录音就绰绰有余了。但是从我们的经验来看，进行录音的访问数量越多，监督效果越好。
3. 监督员在听取访问录音的时候，他们应当使用统一的表格

和评估方法,这一点和电话访谈的监听一样。应当对监督员进行特定的培训,教他们使用同样的评估尺度和标准;他们应当对访问员各个方面的表现都作出评价,要关注优点和缺点两个方面。

4. 通过录音回放进行监督的一个最大问题在于如何对访问员的表现作出及时的反馈。如果反馈不及时,访问员就很难记起访问当时是怎样的情景,从而使反馈的效果大打折扣。但是,即使不能做到及时的反馈,只要对访问的过程进行监督,访问员就会自然而然地对访问的质量提高重视程度。

关于访问员监督效果的一项研究

正如上文所述,对访问员进行系统监督的作用已经引起我们多年的关注。因此,在我们自己的工作中,督导已经成为一项标准的工作。但是,进行这样的监督需要考虑成本问题,而且其他许多调查机构尚未采用这样的做法。我们一直想要证明这样的监督对于访问的质量而言究竟有多大的影响。据我们所知,在已发表的研究中只有一项是研究这个问题的(Billiet 和 Loosveldt,1988)。其中的一些证据表明,访问的监督的确有积极的作用。本书第 7 章介绍的关于访问员培训作用的研究,也有一部分涉及访问监督的作用。

我们可以回顾那个案例。57 名访问员被分为 4 个小组接受培训,培训时间从半天到 10 天不等。培训结束之后,这些访问员被系统地分为三个小组,分别接受监督。

每个访问员都要与一名督导员进行每周一次的电话交流。这时监督员要对访问员在此前一周的表现作出评价和反馈。不同的监督小组之间的区别在于反馈的形式不一样。

第一层次的督导对访问员的反馈只包括工作效率、成本、工时以及受访者接受访问的比率。我们感觉,各个调查机构大致都会对访问员表现中的这几个方面进行评估,也有许多调查机构仅仅将上述这几个方面列为评估范围。

第二层次的督导对访问员的反馈除了上述几个方面之外,还对访问的整体完成情况进行评估。访问的整体评估主要关注访

问员是否遵守了相关的规程，且整体评估的录音也更为完整和精确。虽然这样的评估同样没有直接涉及问一答流程，但是同第一层次的督导相比，这一层次的监督更加关注访问所采集到的数据。此外，这个项目的督导能够复制在方法论上比较严格的调查机构培训自己的访问工作人员的一些反馈。

第三层次的督导要求访问员对所有的访问都进行录音。每周至少要对一个完整录音进行系统评估，监督员所作出的反馈中要包含对访问进行情况的评估。同样，第一层和第二层督导所涉及的内容在这里也都包含在内。

可惜的是，我们无法对这些监督方式加之于访问员表现的影响作出直接的评估，因为我们缺少一、二两组访问员的信息。但是，通过回访以及访问员自己所提供的信息，我们可以对监督的作用进行评估。此外，最重要的是，我们可以将不同小组的访问员所采集的数据进行对比。

需要额外指出的一点是，真正找出监督对于访问员的影响是十分复杂的，因为访问员之前所接受的培训的差异也会影响到监督的效果。举例而言，如果通过录音进行监督，那么培训时间最短的那些访问员会遇到一些比较特殊的问题。

与前面一样，我们将对监督的作用的评估分为四个部分：

1. 访问员的工作态度；
2. 受访者的回应；
3. 访问技巧；
4. 数据质量。

访问员对工作岗位的态度

首先我们要问的是，访问员是否会意识到不同监督小组所作的反馈内容之间的差异，回答是肯定的。和预期的情况一样，访问员对于不同反馈内容的评价还是有差异的，最明显的差异发生在只获得有关成功率和成本之反馈的访问员（第一层）与同时获得关于访问整体质量之反馈的访问员（第二、三层次）之间。

在各个小组就如何进行访问的各种信息来源的重要性进行

评分时，他们对于工作实践、先前经验和培训的各项评分，均高于对督导反馈的评分。尽管如此，各小组对监督重要性的评价还是同我们所预期的一致：在就如何进行访问的各种信息来源的重要性进行评分时，接受第一层次督导的访问员所给出的评分低于其他两组。

对访问员进行录音监督的一个主要理由，是要告诉访问员，他们处理问一答过程的方式是具有极其重要的意义。通过对访问员实际工作的观察，我们得出结论，如果访问员没有受到直接的监督，他们对于标准化的重视程度就会降低。我们认为，对访问的整体表现进行评估和反馈，虽然不一定能了解到数据的采集过程，但至少能让访问员对数据的质量引起重视，从而对整个访问的效果产生积极的影响。

与这一议题相关的资料来源于调查研究中心以前的访问员对于访问员表现的认识。如表8.2所示，分属三个监督小组的访问员之间表现出了显著的差异。虽然各组成员都将获得准确的回答视为最重要的目标，但是第一小组的成员相对于其他两组而言，对受访比例以及效率的重视程度要高得多。相反，第二、三组成员则将“标准化”视为最重要的因素。总的来看，这些数据清晰地表明，对于访问员而言，反馈包含着哪些内容也就意味着调查机构对这些方面更为重视；也就是说，调查机构的关注点是通过这种监督反馈传达给访问员的。访问员的表现中不被评估的那些部分是很难引起他们重视的。这一点与我们所预想的情形一样。

表8.2 访问员就各种目标对于中心工作人员的重要性进行评分：依据监督层次分组

	重要性评分(1—最重要,5—最不重要)		
	监督层次		
目标	Ⅰ	Ⅱ	Ⅲ
回收率	2.5	3.3	3.0*
效率	2.9	3.2	3.2*
标准化	2.6	2.4	2.3*
获得准确的答案	1.5	1.5	1.4
迅速完成所分配的任务	4.6	4.7	4.8

注：* 接受不同监督层次的访问员之间，对目标的重要性的评分存在差异，$p<.05$，F 检验。

最后一个问题是，监督是否会对访问员对于自身工作的感受产生影响。很多人会担心访问员不愿意被录音，不喜欢受到密切的监督。虽然一开始会有一小部分人确实有抱怨和排斥心理，但是一旦我们将监督纳入常规工作，访问员就会逐渐习惯起来。我们从培训和监督中所得的数据与我们的经验是一致的：监督几乎不会使访问员对自身工作产生任何负面情绪。具体而言，访问员所受到的监督与他们对访问工作的感受和评价之间没有任何关联。访问员对于调查项目管理和运作水平的评价倒是与监督的水平相关，但是结果比较奇怪，第二组的评价最低，而第三组的评价最高。

当我们要求访问员对各自所受到的监督打分的时候，我们发现监督较为全面的一组所作的评价显然要更高。此外，受到录音监督的访问员对自己所受的培训的评价也更高。我们认为，这是因为培训是在真正进行访问之前，只有第三组的监督包含了对访问员访问技巧的评估，因此也只有这一组的访问员真正感受到了培训所授技能的作用。总的来说，录音没有任何负面影响，就访问员对访问工作的评价而言，录音有利无弊。

录音对受访者的影响

录音对受访者可能会有两方面不同的影响。一方面，他们可能不愿接受录音，因此对访问表现消极。另一方面，由于录音会对访问员的表现产生影响，因此受访者对访问的态度和看法也会因此而改变。

我们仔细阅读了许多受访者的报告，这些报告说明了他们自己在访问中的应对以及他们对访问员之态度的反应。例如，我们向受访者询问，他们认为访问的重要性如何，他们对其有多大兴趣；我们还向他们询问，是否认为某些问题因为涉及隐私或是难以启齿而影响到回答的精确性。对于这些问题的回答全部都与访问是否被录音有关。

我们同样研究了受访者对访问员表现的评分。虽然我们不指望这些受访者能对访问员的表现给出多少有用的信息，但我们

仍得到了一个有趣的发现。那就是在许多问题上，受访者对访问员的评价大同小异，比如访问节奏的快慢，访问员是直奔主题还是先聊聊其他话题，访问员表现得有多友善或者多专业。对于所有这些评价，受访者的评分与访问是否被录音之间都没有什么具有统计学意义的关联。

我们还让受访者为访问员对回答精确性的重视程度进行打分。同样，尽管我们分析了 1 300 多位受访者的回答，数据的可靠性足够高，但我们仍未能发现这些结果与访问是否被录音之间有任何关系。

在各个监督级别所对应的受访者评价中，唯一一个具有统计学意义的差别，是访问员所表现出来的紧张程度或放松程度。按照受访者的评分，那些被录音的访问员要显得更加紧张一些。此外，值得一提的是，培训时间少于一天且被录音的访问员表现得尤为紧张。

总的来说，从对受访者的回访来看，没有证据说明录音有任何负面作用。当然，我们只有在征得受访者同意的情况下才会录音；只要受访者拒绝，访问员就不会进行录音。不过，受访者拒绝录音的比例不超过 10%。从这一角度来看，录音显然是用来控制访问质量的一个可行方法。

监督对于访问技巧的影响

之前我们说过，我们无法对不同的监督小组成员的实际访问过程进行对比，原因是如果不进行录音，我们就不知道访问过程究竟是怎样的。但是，通过对录音下来的那一部分访问所进行的分析，我们发现，系统的反馈通过长期的影响能够对访问员的访问技能产生积极的影响。我们将每个访问员所分得的样本随机地分成两部分，进而对这两部分所获得的数据结果进行对比，分析其中的变化。对于被录音的那些访问员，我们将比较两部分样本中的受访者对访问员表现的评分。

考察表 8.3 的时候应当注意三点。首先，我们所观察的是一个非常短的时间段内的变化。这项研究中访问员的工作只持续 5 ~6周。因此，我们只能将访问员前 3 周的表现与后 3 周的表现

进行对比。第二，虽然我们要求访问员对所有的访问都进行录音，但我们每周只选取其中一到二个访问进行讨论。他们每周会与各自的监督员进行一次交谈，讨论本周的表现。因此，访问员总共与监督员只进行 3～4 次反馈交流。最后，我们不会真的淘汰和辞退任何访问员。监督的目的只是为了让访问员对访问的标准化有更好的认识，并由此观察访问员的实际表现状况。

表 8.3 依据访问录音编码对访问员行为进行评估：前一半任务和后一半任务比较（包括各个小组，只接受了第三层次督导）

	任务分类		
访问员行为的录音编码	前一半任务	后一半任务	p
不正确追问/访问的平均数	14	12	n.s.
诱导性提问/访问的平均数	6	5	<.05
对答案追问/访问不充分次数的平均数	6	6	n.s.
记录封闭式问题答案/访问不正确的平均数	1	1	n.s.
记录开放式问题答案/访问不正确的平均数	3	3	n.s.
不合适反馈/访问事件的平均数	1	1	n.s.
			n.s.
被评为优秀或者满意的访问所占百分比			n.s.
逐字逐句提问	63%	70%	n.s.
追问封闭式问题	61%	73%	<.03
追问开放式问题	42%	48%	n.s.
记录封闭式问题的答案	88%	91%	n.s.
记录开放式问题的答案	68%	74%	n.s.
非偏误性人际行为	81%	87%	n.s.

出乎我们意料的是，随着调查的进展，我们没有发现访问员的访问技能有多大的进步。虽然我们所观察到的差异与预期的情形大致相同，但受访者对于访问员前后两部分的评分是基本相同的。此外，对培训时间不同的各个小组的比较，所得到的结果也大致相同。培训时间少于 1 天的小组在访问刚开始时的访问技能非常欠缺，而通过录音和评估反馈，他们的访问技能也几乎没有得到任何提高（见表 8.4）。唯一一个真正显示出效果的是培训时间为 5 天的小组，他们在后半部分的表现有明显的提高。

表 8.4 依据前一半任务编码和后一半任务编码对访问员行为评分:不同培训时间比较(只包括接受第三层次督导的访问)

在逐字逐句提问方面被评为优秀或者满意的百分比

	任务分类		
培训时间的长短	前一半任务	后一半任务	净变化
<1 天	32%	28%	-4%
2 天	83%	84%	+1
5 天	64%	79%	+15
10 天	85%	83%	-2

在追问开放性问题方面被评为优秀或者满意的百分比

	任务分类		
培训时间的长短	前一半任务	后一半任务	净变化
<1 天	18%	13%	-5%
2 天	57%	54%	-3
5 天	44%	61%	+17
10 天	70%	69%	-1

这些数据表明,监督和反馈本身不会提高访问的技能水平。如果没有进一步的培训,那么,访问员在培训结束时所表现出来的能力水平也就是他们所能达到的最高水平了。我们的结论是,包含指导练习的培训是提高标准化访问技能的关键所在,而监督的关键作用则在于确保这些技能得到正确的运用。

监督对数据质量的影响

我们使用两种方式来测量访问员所采集到的数据质量:一是 rho,即组内相关系数,用于测量访问员的一致性;二是测量偏差,用来衡量访问员所获得的数据与平均值相比偏差的大小。表 8.5 列出了相关数据。

表 8.5 各项的组内相关系数以及对可能属于系统偏差的问题的平均标准分(×1 000):不同督导层次的比较

	督导层次		
数据质量指标	Ⅰ	Ⅱ	Ⅲ
组内相关系数	.012	.010	.008*
标准分(偏差)(包括各个培训组)	4	5	20**
标准分(偏差)(排除了培训时间少于 1 天的组别)	0	4	34***

注: * $t = 1.6$,第三层次和第一、二层次比较,$p < .05$,单尾检验。分数越低说明访问员效应越小。
** $t = 1.0$,第三层次和第一、二层次比较,p 不显著。分数越高说明偏差越小。
*** $t = 1.8$,第三层次和第一、二层次比较,$p < .04$,单尾检验。分数越高说明偏差越小。

随着督导层次的提高,rho 具有细微但显著而单纯的增加;两个加强监督的计划,对于从总体上减少组内相关系数具有显著作用。而且,如果我们只将那些被录音的访问员与其他访问员相比,这种差异具有统计学上的显著性,符合我们的预期(单尾检验)。

但是,在这个特殊的案例研究中,访问员接受的培训和督导在减少组内相关系数方面具有交互效应。特别是,更高强度的督导对于那些只接受了很少培训和接受了最多培训的访问员来说,具有显著的帮助;而对于那些只接受了 2 天到 5 天培训的访问员而言,在这项关于访问员的研究中,这种监督没有显著的作用。

在录音记录中,偏差的变化趋势与我们的预期相符合,但是,这种趋势没有统计学显著意义上的正效应。此外,这也与培训具有显著的交互效应。那些只接受了少于 1 天的培训、但做了录音记录的访问员,其数据的偏差最大。如果这部分访问员被排除在外,则资料显示,录音记录对于减少调查估计的偏差具有显著的正效应。

结 论

我们相信,访问监督的质量是决定访问员的表现以及所采集

的数据质量的一个关键因素,尽管监督本身并不能保证访问和数据的质量。一个意外的发现是,监督并不能弥补培训时的不足。如果通过培训,访问员仍未能掌握访问的标准和技能,那么依靠监督时的评估和反馈也不能使他们有所提高。事实上,甚至有证据表明,培训不够充分的访问员在密切的监督之下,其表现反而比不受监督时更差,尤其是在偏差方面。

我们的研究得出的另一个重要的结论是,对于那些对自己的访问技能抱有充分自信的访问员而言,监督是非常重要的。研究数据表明,培训时间超过 10 天的访问员,他们虽然对访问的规程和技能十分熟悉,但是如果不进行录音,所得的数据质量要远远低于我们的预期,甚至还不如那些培训时间较短的访问员。毋庸赘言,培训时间较长且接受录音的访问员的表现是最符合标准化要求的。我们相信,资历较深的访问员,由于受过反复的培训,在访问中自作主张的情况会越来越多,只有系统的监督才能控制这一现象。

简而言之,我们认为,对访问员在访问中的表现进行密切的监督虽然不能教给他们新的访问技能,但能确保他们正确运用在此前的培训中所学到的技能。虽然我们的数据来源只是一项为期 6 周的研究,但其他人的研究结论也同样支持了我们的结论。例如,Cannell 等进行了一项为期 5 周的调查,目的是采集人们的住院情况信息。研究发现,随着调查的进展,访问与最初培训所间隔的时间越长,所采集的数据质量也就越差。他们还发现,一个访问员所进行访问的次数同采集到的数据质量成反比。Bradburn 和 Sudman(1979)所做的另一项研究也表明,资历较深的访问员同新手相比,在提问时要明显表现得更为随意(错读或随意添加和删减字词)。简单来说,虽然我们的数据还不够充分,但确有迹象表明,随着访问员对自己的访问能力越来越自信,他们会更容易偏离标准化的要求。

随着电话访问越来越普及,对访问进行系统监督的可行性也越来越大。事实上,电话访问的一个潜在优势就是,它能够保证访问的标准化程度。此外,对电话访问进行监督的优势在于能够及时地向访问员作出反馈。

我们已经证明录音是确保面对面访问质量的一个有效方法。

证据表明,多数受访者不介意被录音,录音不会改变受访者对待访问的态度;其次,有明确的数据表明,通过录音监督,访问员对于标准化的重视程度也会更高;最后,虽然录音和访问复核的作用并不总是那么明显,但是除了培训时间少于 1 天的小组以外,录音对于数据质量的积极作用还是得到证实的。这表明,录音监督对于高质量的数据采集工作而言,是十分重要的一个环节。

第9章 如何减少调查中的访问员相关误差

How to Reduce Interviewer - Related Error in Surveys

本章我们将对本书所表述的内容进行总结，并围绕如何使访问员相关误差最小化提供综合性的建议。

如果访问员使用各自不同的方法处理访问过程，那么就会使调查产生误差。在调查中，受访者所给出的回答其实就是这个测量过程所得到的产品。如果各个访问员处理问题的方式各不相同，或是对待不同受访者的方式各有差异，那么就会对所采集到的回答产生相应的影响，进而降低测量的精确性和可靠性。研究者要测量的事物，例如事件发生的概率、真实的分数等，这些答案的误差在很大程度上可以归因于访问员。

在调查中，访问员所实施的数据采集工作，是研究者之工作的延伸。要使访问员的人为误差降到最低，研究人员就必须确保所有的访问员在面对所有的受访者时，使用一致的方法来处理数据采集流程。研究人员可以通过以下6种方法对访问的一致性施加影响：

1. 问题的遣词造句；
2. 向访问员提供具体的指示或手册，教会他们如何处理问—答过程；
3. 向访问员提供具体的指示或者程序，告诉他们如何建立并维护与受访者的关系；
4. 选择访问员；
5. 培训访问员；
6. 在数据采集过程中监督访问员。

问题设计

我们在别处已有提及(Fowler,1984),为了进行可靠测量,问题的拟写必须至少符合三个标准:

1. 问题的文字表述全面,能够向受访者提供回答问题所需的全部信息;
2. 问题对每个受访者都有同样的含义;
3. 要让所有受访者知道什么类型的回答才是恰当的回答。

研究者使调查标准化的一个关键贡献就是拟定出满足上述三个标准的问题。如果研究者设计的问题不满足上述任何一个标准,都将给受访者和访问员制造麻烦,而这些麻烦通常是很难用标准化方法进行解决的。

设计的问题要能让所有的受访者都完全清楚,并且让他们的理解保持一致,这是一件很艰难的事情。但是,研究者可以做得比现在更好。常用的预调查对于识别此类问题还是远远不够的。从我们的经验来看,经验丰富的访问员如果在预调查中遇到需要重新表述问题的情况,通常不会太注意。我们认为,如果在预调查中对问题的设计和编写提高重视程度,一定会得到事半功倍的效果。我们在调查中应当更加广泛地使用以下三个方法:

1. 在实验室条件下运用认知心理学的技术来逐一测试回答任务以及对关键概念的理解,在调查领域这种做法现在才开始流行;
2. 更多地使用焦点讨论小组的方法,了解人们对调查所涉及的措辞、概念以及问题是如何理解的;
3. 对预调查进行录音和编码,从而提供系统的信息,鉴别那些读起来比较拗口、需要进一步说明、需要反复追问,以及由于其他原因会给受访者和访问员制造麻烦的问题。

然而,即使运用了这些方法也不一定能确保问卷的完美无缺。但是我们相信,很多问题都不符合前文所述的三个标准,即便是一些专业机构所进行的大型调查也不例外。虽然上述这三

种方法并不一定能将所有的问题一网打尽,但其效果是毋庸置疑的。我们认为,没有哪种方法能够比改善问卷质量更有助于提高访问的标准化程度。

访问流程

我们认为,向访问员提供标准化的说明,指导他们处理问—答过程,这种方法是十分有效的。以下是四个标准规则:

1. 严格按照问卷书面的表述提问;
2. 用非引导性的方法对不完整的回答进行追问;
3. 记录答案时不能加入访问员的判断;
4. 与受访者保持一种中立的、非评判性的关系。

如果这些原则得到执行,情况就会好很多。但是,在遇到这四个原则也无法对付的问题时,标准化就会受到影响。

我们已经说过,处理此类问题的最首要、最好的方法就是在设计问题的时候尽量避免这种情况的发生。但是,这种情况的确无法避免,一旦访问员遇到类似的问题,就容易违反标准化的规则,用各自的方法来解决问题。实际上,当一个问题对于受访者而言表意不清或者根本不适用时,访问员就容易将问题加以改动或重新表述,来迎合受访者的实际情况。如果访问员当时没有其他工具,或者不知道如何将访问进行下去的时候,他们就更容易犯这样的错误。上述四条规则并没有告诉访问员在遇到这种情况,访问中断时如何将访问继续下去。

当标准化访问因受访者表现不当而中断的时候,访问员的正确做法是向受访者就标准化测量作出解释,并让他们知道在访问中出现问题是正常的情况。如果不加以解释,受访者自己是不会明白为什么在不清楚问题含义的情况下还要回答问题,为什么这样的回答同样有助于调查的测量。因此,研究人员需要对访问员进行培训,指导他们按照统一的模式对相关问题作出解释,引导受访者回答问题,从而保证整个问答过程按照标准化的要求进行。

与受访者之间的关系

很少有调查机构会就如何与受访者建立关系方面给予访问员太多的指导，这是因为很难说清什么样的关系对于标准化访问才是最好的。在第 4 章，我们曾经深入考察过问—答关系的两个方面：一是受访者从访问员那里得到的关于此项调查的信息；二是访问员与受访者进行交流的风格。我们的结论是，这两方面都无法实现标准化。在一次访问中，不同的受访者所提供的信息量以及他们接受访问的目的是有很大差异的。鉴于大多数调查所使用的方法，受访者所提供的信息量也是很少有上升空间的。所以，受访者对于一项调查及其调查目的的了解程度以及他们各自接受访问的目的肯定是各不相同的。

至于问答双方之间所建立的关系，虽然大家都知道理想的状态是一种友善而专业的关系，但在实际访问中，双方所建立的关系是各种各样的。要想让所有的问答组合都建立起某种理想化的关系模式，显然是不切实际的。

虽然受访者对访问员的态度，以及双方所建立之关系的性质是不能被操纵和控制的，但是，可以将不同访问员和受访者之间的回应表现统一起来。Cannell 及其助手提出了五种方式，从而让不同的访问员和受访者之间的沟通保持一致性：

1. 访问节奏；
2. 向受访者提供示范；
3. 无论表现如何，在访问开始之前和访问过程中，向受访者提供统一的说明；
4. 对符合预期的行为进行模式化的鼓励；
5. 要求受访者作出承诺，保证自己的表现能够达到某一标准。

访问员在访问过程中通过使用上述策略，能够避免不同的访问员和受访者之间出现不一致的情况，从而提高数据的质量。

访问员的选择

在选择访问员的时候应当考虑两类特征因素:一类特征会影响到访问员驾驭调查过程的能力;另一类性格特征则会对访问的情境产生微妙的影响。

在访问员的能力方面,除了良好的读写技巧和相对亲善的性格之外,并没有其他更特殊的筛选标准。

当然,访问员的很多人口学特征的确会对访问的情境产生影响,比如年龄、性别、阶层以及种族和族群背景。对于大多数调查的大多数主题以及大多数受访者而言,访问员的这些人口学特征并不会对受访者的回答产生任何影响。并且一般而言,我们认为之前各章所列出的这些访问流程,以及访问员所接受的培训和监督,可以帮助克服这些人口学特征对访问员相关误差可能带来的影响。

但是,如果问题的内容与访问员的个体特征直接相关,以至于受访者可以判断(或者假想)出访问员本人对于某个答案的偏好,那么,访问员对所采集的数据的影响就很明显了。例如,受访者在面对具备犹太人外貌或名字特征的访问员时,所表现出的反犹太倾向要相对微弱一些。虽然对于这种问题而言很难说哪些回答是最好的或是最精确的,但很明显的一点是,如果访问员的个体特征与问题的回答直接相关,那么就有可能使测量过程产生误差,导致数据的不可靠性。

遇到这种问题时,研究人员可以考虑两种解决方法。第一,如果他们认为访问员的某方面特征会对受访者的回答产生影响,那么可以挑选在这方面具备相同特征的人作为访问员;第二,如果不能确保所有的访问员都具备相同的个体特征,那么可以在分配任务的时候,将代表性样本或者配对样本分配给不同特征的访问员(如男性和女性访问员)。这样,在分析数据的时候,就便于将这些个体特征(如性别)所造成的影响考虑在内。

但总而言之,我们在这方面得出的一个最重要的结论是,对于绝大多数调查而言,挑选访问员对于控制访问员相关误差而言并不是一个重要策略。

访问员的培训

即使是聪明人也并不会自然而然地成为一个标准化访问的访问员。要成为一名合格的标准化访问员，至少要接受 2～3 天的培训，培训的内容包括对各个访问流程的指导练习以及处理问题的技巧。

在我们为访问员培训所设计的研究中，有一组成员只接受了为期半天的培训，内容只包括讲座和示范。此外，我们发给每个访问员一本专业的指导手册，在调查正式开始之前让每个组员做了一次访问练习。在调查开始后，再对他们的工作进行监督，每周由经验丰富的督导与他们进行交流和反馈。因此，虽然这组成员所受的培训十分简单，但是他们的工作情境是十分专业的，而且他们已经明确无误地被告知什么样的做法才是标准合格的做法。尽管如此，这组访问员的标准化访问技能还是普遍欠缺的，即便同培训时间只有 2 天的小组相比也要逊色很多。虽然在获得合作、与受访者建立良好的关系以及在效率和效能方面，他们与其他小组的访问员表现相当，但访问的录音显示，在提问、非引导性追问、开放性问题的答案记录以及在交流过程中保持中立这些方面，他们的表现还是十分欠缺的。此外，即使在这些问题上督导对他们作出了反馈和建议，至少在我们的研究所持续的 6 周时间内，他们在这些技巧上并未表现出显著的提高。

以上所提及的这些关键技能还仅仅是数据采集的一部分要求。但是已经有足够的证据表明，培训时间最短，即只进行了大约半天的小组，在访问技巧方面所接受的指导是十分不足的。为了确保数据采集的一致性和高质量，必须对访问员进行更多的培训。

督　导

我们相信，对访问过程进行监督是确保调查标准化的一个必要条件。对于电话访问而言，监督是指对访问员处理访问过程的方式以及他们与受访者之间建立起的关系进行统一的监听、评估和反馈。对于面对面的访问而言，可以通过录音对访问过程进行

评估和反馈。我们的调查结果显示,这种方法有助于减少由访问员所引起的误差。此外,对于培训时间超过一天的访问员,监督也有助于减少数据的偏差。对访问过程的整体评估和反馈可以让访问员认识到他们采集数据的方法的重要性。而对访问的直接监督更能提高他们对标准化的重视程度,从而进一步提高他们所采集到的数据的质量。因此我们认为,这样的监督应当被列为一项调查的标准流程之一。

结　论

在过去的 10 年中,总体调查设计(total survey design)概念已经成为方法论学者的流行趋势。这里所说的方法学家的工作是综合考虑数据采集过程的各个方面,对调查项目进行资源配置。

很多年来,衡量调查质量的指标似乎只有两个:抽样设计和受访者接受访问的比例。即使在今天,人们如果在一项调查的结果之后看到这样的注释:“误差在 ±4% 之间”,这样的标注所反映的也仅仅只是由于抽样所带来的误差。也就是说,这一误差产生的原因在于,数据的来源不是总体内部的每一个人,而只是其中具有代表性的一部分。这个误差的数字是很容易计算的:它只是计算了可以观察到的一部分误差,来源于样本规模和抽样设计。但是,这样的注释数据将其他可能引起数据误差的因素完全忽略掉了,因为它默认了抽样是一项调查的唯一误差来源。

研究者之所以将抽样误差视为调查中的主要的、甚至唯一的误差来源,是因为抽样误差是很容易进行观察和计算的。调查中另一个便于观察和计算的数据则是样本接受访问的比例。但是,由处理访问过程的方式所导致的误差(通常称为响应误差(response error)),一般来说不易观察,因此常常被忽略。

研究人员在进行调查设计和配置资源的时候,目标都在于尽量减少调查中的误差。他们通常所提出的方法都是扩大样本容量,或是更加精确合理地设计样本。另一个较为常用的方法就是采取措施使得样本接受访问的比例尽量提高。但是,很少有人提及本书所关注的这些方面。但是,我们有理由相信,本书所列的这些方法与扩大样本容量相比,能够更加有效地提高调查数据的

质量。此外,从节省成本的角度考虑,这些方法中绝大部分相对而言都是十分便宜的。

作为总结,表 9.1 为研究人员列出了一些提高数据质量的可行方法,包括扩大样本容量以及其他一些减少访问员相关误差的方法。从我们现有的数据来看,通过减少访问员人为误差的方法来提高数据质量要比扩大样本容量更有利于节省成本。而且,几乎可以肯定,表格中的数据还没有充分反映出这些方法对于提高数据质量的作用。例如,在设计调查时提高问题的质量是控制访问员误差的最佳途径。此外,高质量的问题也可以使受访者对问题的理解和回答更加准确。

我们在第 2 章中提到过,访问的质量和调查的设计之所以得不到重视,是因为访问过程中产生的误差和受访者所引起的误差只有通过特别的方法才能得到观察。如果研究人员不愿意去控制这种访问员误差,那么自然就不会花大力气对此进行测量。但是,随着信息采集的重任越来越多地落到社会科学家的身上,他们也日益感受到科学测量和减小误差的迫切性。仅仅承认问卷的设计是一门艺术还是远远不够的。艺术意味着除了个人的品味之外,没有其他评判优劣的标准,而问卷设计显然不是这样。我们已经为问卷设计总结出几条十分具体的标准,其中最核心的一条就是所提出的问题一定要便于使用标准化的统一方法来处理和回答。

同样,要当好一名访问员仅仅靠喜好和感觉也是远远不够的。访问员的好与坏都有具体的评判标准,他们必须遵循一定的规则。访问员处理访问的方式是十分重要的。我们已经列出了许多标准和原则,要确保访问的质量和测量过程的科学性,访问员就必须遵循这些原则。此外,我们也为研究人员指出了一些监督访问员实际表现的方法,用以考察他们是否按规定的要求做了,是否能得出优质的测量结果。

我们的知识仍然存在一些不足。我们相信,我们将来一定能找到更好的办法来测试调查问题的质量。我们认为,我们甚至可以研究出受访者接受调查的最好方式。我们也很可能会提出更好的规则和流程,以帮助访问员更好地对受访者进行指导。但是,我们相信,我们在本书中所列出的这些方法,已经得到了经验

数据的支持,完全足以使访问和数据的质量得到大幅的提升。我们认为,调查研究领域目前所存在的问题并不是缺少有关调查方法的知识,而在于研究者对提高访问及数据质量缺乏足够的重视。我们由衷地希望本书能有助于这一现象的改善。

表 9.1　降低估计标准误差的五种方法

策略	方式	可能提高成本	对标准误差的效应
样本规模[1]	提高有效样本规模 20%	在数据采集和资料简化方面的成本会同比例提高	降低 10%
访问员培训[2]	访问员接受的基本培训从不足 1 天提高至 1 天或者 2 天	每个访问员额外增加的培训,每天大约相当于一个访问员 12 小时的工资	对最受访问员影响的 1/3 的调查项而言,降低 10%
录音督导[3]	对访问进行全体录音或者抽样录音,每周对每位访问员进行评价,提供反馈	每个访问员每周花费 2 小时	对最受访问员影响的 1/3 的调查项而言,降低超过 10%
问题设计[4]	重写问题,减少追问,使管理和提问更加容易	依据试调查录音将访问员行为进行分类,耗时大约是访问时间的两倍,此外加上重编问题的时间	成效尚未得到证明,但是数据质量将获得显著改善
每个访问员访问次数[5]	访问员增加 25%,每人减少 20% 的工作量	很难估计,但是毫无疑问增加的成本将小于为了获得同样效果而进行的样本规模改变	对最受访问员影响的调查项而言,降低 10%

注:1. 如果使用复杂的抽样设计,而非简单随机抽样,为了达到同样的效果,这一方法将增加超过 20% 的访问量。

2. 很显然,在最基本培训的基础上加强培训,将对减少标准误差产生直接效应。但是,更多的培训也将有利于减少资料偏误。

3. 如果在没有录音和反馈之时访问员表现差劲,投入时间更多也许会获得更好的效果。同时,如果给予访问员充分的培训,偏误也会显著降低。

4. 也可能有助于受访者的回答。

5. 访问员精疲力竭的情况减少,同时也可以减少误差。



Anderson, B. A., B. D. Silver. and P. Abramson. 1988a. "The Effects of the Race of Interviewer on Race-Related Attitudes of Black Respondents." *Public Opinion Quarterly* 52(3)289-324.

Anderson, B. A., B. D. Silver, and P. Abramson. 1988b. "The Effects of Race of the Interviewer on Measures of Electoral Participation by Blacks." *Public Opinion Quarterly* 25(1)53-83.

Billiet, J. and G. Loosveldt. 1988. "Interviewer Training and Quality of Responses." *Public Opinion Quarterly* 52(2)190-211.

Bradburn, N. A., S. Sudman, et al. 1979. *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.

Cannell, C. F. and F. J. Fowler. 1964. "A note on Interviewer Effect in Self-Enumerative Procedures." *American Sociological Review*(29):276.

Cannell, C. F. and F. J. Fowler. 1965a. "Comparison of Hospitalization Reporting in Three Survey Procedures," *Vital and Health Statistics* Series 2, (8), Washington, DC: U. S. Government Printing Office.

Cannell, C. F., F. J. Fowler, and K. H. Marquis. 1965b. *Respondents Talk About the National Health Survey Interview*. Survey Research Center. Ann Arbor: Universitv of Michigan. Mimeographed.

Cannell, C. F., F. J. Fowler, and K. H. Marquis. 1965c. *Report on Development of Brochures for H. I. S. Respondents*. Survey Research Center. Ann Arbor: University of Michigan. Unpublished manuscript.

Cannell, C. F. , G. Fisher, and Thomas Bakker. 1965d. "Reporting of Hospitalization in the Health Interview Survey, " *Vital and Health Statistic. Series* 2 (6). *s*Washington, DC: U. S. Government Printing Office.

Cannell, C. F. , F. J. Fowler, and K. H. Marquis. 1968. "The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews, " *Vital and Health Statistics*, Series 2(26). Washington, DC:U. S. Government Printing Office.

Cannell, C. F. , K. H. Marquis, and A. Laurent. 1977a. "A Summary of Studies. " *Vital and Health Statistics*, Series 2(69). Washington, DC:U. S. Government Printing Office.

Cannell, C. F. , L. Oksenberg, and J. M. Converse. 1977b. *Experiments in Interviewing Techniques:Field Experiments in Health Reporting*:1971-1977. Hyattsville. MD:NCHSR.

Cannell, C. F. et al. 1987. "An Experimental Comparison of Telephone and Personal Health Interview Surveys. " *Vital and Health Statistics*, Series 2 (106). Washington, DC: U. S. Government Printing Office.

Cannell, C. F. and L. Oksenberg. 1988. "Observation of Behaviour in Telephone Interviewers. " Pp. 475-495 in *Telephone Survey Methodology*, edited by R. Groves et a1. New York:John Wiley.

Cohen, A. R. 1958. "Upward Communication in Experimentally Created Hierarchies. "*Human Relations* 11:41-53.

Converse, J. M. and S. Presser. 1986. *Survey questions*. Beverly Hills, CA:Sage.

Converse, J. M. 1987. *Survey Research in the United States*. Berkeley:University of California Press.

Cronbach, L. J. and P. E. Meehl. 1955. "Construct Validity in Psychological Tests. " *Psychological Bulletin* 52:281-302.

DeMaio,T. J. , ed. 1983. *Approaches to Developing Questionnaires*. Statistical Policy Working Paper 10. Washington. DC: U. S. Government Policy Office.

Erlich, J. and D. Reisman. 1961. "Age and Authority with Interview." *Public Opinion Quarterly*, 25: 39-56.

Fowler, F. J. 1966. *Education, Interaction and Interview Performance*. Ph. D. dissertation, University of Michigan.

Fowler, F. J. 1984. *Survey Research Methods*. Beverly Hills. CA: Sage.

Fowler, F. J. & Mangione, T. W. 1986. *Reducing Interviewer Effects on Health Survey Data*. Washington, DC: National Center for Health Services Research.

Groves, R. M. and R. L. Kahn. 1979. *Surveys by Telephone*. New York: Academic Press.

Gloves, R. M. and L. J. Magilavy. 1980. "Estimates of Interviewer Variance in Telephone Surveys." Proceedings of the American Statistical Association, Survey Research Methods Section, 622-627.

Guralk, D. B., ed. 1976. *Webster's New World Dictionary*. Cleveland: Collins and Wood Publishers.

Hensen, R. M. 1973. *Effects of Instructions and Verbal Modelling on Health Information Reporting*. Ann Arbor: Survey Research Center. University of Michigan.

Hyman, H. A., J. Feldman, and C. Stember. 1954. *Interviewing in Social Research*. Chicago: University of Chicago Press.

Jabine, Thomas B., Miron L. Straf, Judith M. Tanor, and Roger Tourangeau, eds. 1984. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, DC: National Academic Press.

Kahn R. . and C. F. Cannell. 1958. *Dynamics of Interviewing*. New York: John Wiley.

Kelman, H. C. 1953. "Compliance, Identification, and Internalization: Three Processes of Attitude Change." *Human Relations* 6: 185-214.

Kish, L. 1962. "Studies of Interviewer Variance for Attitudinal Variables." *Journal of the American Statistical Association* 57: 92-

115.

Kish, L. 1965. *Survey Sampling*. New York: Wiley.

Locander, W. ,S. Sudman,and N. Bradburn. 1976. "An Investigation ofInterview Method, Threat and Response Distortion. " *Journal of the American Statistical Association*. 71: 269- 275.

Lundberg,G. A. ,et al. 1949. "Attraction Patterns in a University. " *Sociometry*. 12: 158-159.

Marquis, K. H. ,C. F. Cannell, and A. Laurent. 1972. "Reporting Health Events in Households Interviews: Effects of Reinforcement, Question Length, and Reinterviews. " *Vital and Health Statistics*, Series 2 (45). Washington, DC: U. S. Government Printing Office.

McKinlay, S. M. , D. M. Kipp, P. Johnson, K. Downey, and R. A. Carleton. 1982. "A Field Approach for Obtaining Physiologic Measures in Surveys of General Populations. " Proceedings of the fourth conference on health survey research methods. Washington. DC: National Center for Health Services Research. 195-204.

Meisel. A. and H. G. Roth 1983. "Toward Informed Discussion of Informed Consent: A Review and Critique of the Empirical Studies. " *Arizona Law Review* 25: 266-341.

Mishler, E. G. 1986. *Research Interviewing*. Cambridge. MA: Harvard University Press.

Newcomb,T. M. 1961. *The Acquaintance Process*. New York: Holt, Rinehart,& Winston.

Payne, S. L. 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.

Robinson, D. , and S. Rohde. 1946. "Two Experiments with an Anti-Semitism Poll. " *Journal of Abnormal Social Psychology* 41: 136-144.

Sanders, Barkev S. 1962. *A Health Study in Kit Carson County*. Public Health Service Publication, No. 844. Washington, DC: U. S. Government Printing Office.

Schuman, H. and J. M. Converse. 1971. "The Effects of Black and

White Interviewers on Black Responses in 1968." *Public Opinion Quarterly* 35: 44-68.

Schuman, H. and S. Presser. 1981. *Questions and Answers in Attitude Surveys*. New York: Acàdemic Press.

Stokes, S. L. 1986a. "Estimation of Interviewer Effects in Complex Surveys with Application to Random Digit Dialing." Proceedings Annual Research Conference, Bureau of the Census. 2: 21-31.

Stokes, S. L. 1986b. "Estimating Interview Variance for Dichotomous Items Using a Latent Variable Model." *Proceedings* of the Meeting of the American Statistical Assocation, Survey Research Section, 278-280.

Stokes, S. L. and M. Yeh. 1988. "Searching for Causes of Interviewer Effects in Telephone Surveys." Pp. 357-373 in *Telephone Survey Methodology*, edited by R. Groves et al. New York: John Wiley.

Sudman, S. and N. M. Bradbum. 1974, *Response Effects in Surveys: A Review and Analysis*. Chicago, Aldine.

Sudman, S., N. M. Bradbum, E. Blair, and C. Stocking. 1977. "Modest Expectations: The Effects of Interviewers' Prior Expectations on Responses." *Sociological Methods and Research* 6: 177-182.

Weiss. C. H. 1968. "Validity of Welfare Mothers' Interview Responses." *Public Opinion Quarterly* 32: 622-633.

译校者后记

这是一部关于调查误差的方法论专著。作者就访问员相关误差(interviewer-related error)的概念、来源及评估方法进行了系统探讨,并以大量实地调查资料和实验数据为基础,从学理层面就优化访问情境、改进问卷设计、选聘合适的访问员、培训访问员和监督访问员对于减少误差、提高数据质量的效应进行了深入研究,并提出了有针对性的建议。

近三十年来,抽样调查技术在中国市场调查、民意调查和学术研究领域得到了非常广泛的应用,为实际工作和学理研究积累了大量数据资料,极大推动了决策工作的科学化和学术研究的实证化。本书的翻译和出版,可以帮助广大实际工作者和理论研究者深化对调查误差的认识,从而为降低调查误差、进一步提高数据质量提供借鉴。

应当指出的是,迄今为止,调查研究方法在中国的发展仍然主要停留在“引进”和“应用”阶段,引进的多为方法教材,专门的研究性著作尚不多见。本书中译本的出版,对于推动调查研究方法由“引进”和“应用”阶段提升到“研究”阶段,以中国经验为基础撰写社会科学方法论方面的专著,提供了一个范本和可供比较的标杆。

本书翻译是集体合作的成果。徐方敏女士提供了全书除图表及部分章节之外的大部分初稿;孙龙对本书的前言和第 1 至第 3 章进行了全文重译,对第 4 至第 9 章进行了补译和逐字逐句校对,翻译制作了全书的所有图表;张小劲教授统校全书并修改定稿;最后,黄英女士仔细阅读了清样,纠正了不少笔误和缺漏。

本书得以出版,还要感谢雷少波先生的努力和敦促!正是雷少波先生和重庆大学出版社在推进万卷方法丛书方面的热情和

投入感动了本书的译校者，促使我们在非常紧张的教学、研究和工作之中，仍然抽时间非常认真地承担了本书的翻译工作。

然而，由于时间、精力及学识所限，译稿之中仍可能出现误读及疏漏之处，尚请广大读者不吝指正，以期将来改进。

译校者谨识

2009 年 4 月 20 日

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝标准化调查访问　如何实现访问员相关误差最小化
作者＝
页数＝１３９
ＳＳ号＝
ＤＸ号＝
出版日期＝
出版社＝